THE ANNALS

of

MATHEMATICAL STATISTICS

FOUNDED AND [illegible]
[illegible]

THE OFFICIAL JOURNAL OF THE INSTITUTE OF MATHEMATICAL STATISTICS

Contents

VOL. 22, No. 4 — [illegible], 1951

THE ANNALS OF MATHEMATICAL STATISTICS

Published quarterly by the Institute of Mathematical Statistics in March, June, September and December at Baltimore, Maryland.

INSTITUTE OF MATHEMATICAL STATISTICS

General Office: Business Administration Building, University of Michigan, Ann Arbor, Michigan
C. H. Fischer, Secretary-Treasurer
This address should be used for all communications concerning membership, subscriptions, changes of address, back numbers, etc., but not for editorial correspondence. Changes in mailing address which are to become effective for a given issue should be reported to the Secretary on or before the 15th of the month preceding the month of that issue.

Editorial Office: Department of Mathematical Statistics, Columbia University, New York 27, New York
T. W. Anderson, Editor
Manuscripts should be submitted to this address; each manuscript should be typewritten, double-spaced with wide margins, and the original copy should be submitted (preferably with one additional copy). Footnotes should be reduced to a minimum and whenever possible replaced by a bibliography at the end of the paper; formulae in footnotes should be avoided. Figures, charts, and diagrams should be professionally drawn on plain white paper or tracing cloth in black India ink twice the size they are to be printed. Authors are requested to keep in mind typographical difficulties of complicated mathematical formulae.
Authors will ordinarily receive only galley proofs. Fifty reprints without covers will be furnished free. Additional reprints and covers furnished at cost.

Subscription Price: 1951: \$10 per year in Western Hemisphere; \$5 elsewhere; single issues \$3; back numbers \$10 per volume, \$3 per issue.
1952: \$12 per year in U. S. and Canada; \$10 elsewhere.

COMPOSED AND PRINTED AT THE

WAVERLY PRESS, INC., BALTIMORE, MARYLAND, U. S. A.

Entered as second-class matter at the Post Office at Baltimore, Maryland, under the act of March 3, 1879.

EXISTENCE OF CONSISTENT ESTIMATES OF THE DIRECTIONAL PARAMETER IN A LINEAR STRUCTURAL RELATION BETWEEN TWO VARIABLES[1]

By Jerzy Neyman

University of California, Berkeley, California

Summary. Let Z_n denote the system of $8n$ independent pairs of measurements (X_{ik}, Y_{ik}), for $i = 1, 2, \cdots, n$ and $k = 1, 2, \cdots, 8$, of two nonobservable random variables ξ_{ik} and η_{ik}, known to satisfy a linear relation of the form $\xi_{ik} \cos \theta^* + \eta_{ik} \sin \theta^* - p = 0$, where p is an arbitrary real number and θ^* may have any value between the limits

$$-\tfrac{1}{2}\pi < \theta^* \leqq \tfrac{1}{2}\pi.$$

The purpose of the paper is to construct a class of estimates $T_n(Z_n)$ of the parameter θ defined as follows: when $\theta^* = \frac{1}{2}\pi$ then $\theta = 0$; otherwise $\theta \equiv \theta^*$. Each estimate $T_n(Z_n)$ of the class considered converges in probability to θ as $n \to \infty$ under the following conditions: (i) except when $\theta = 0$, the variables ξ_{ik} are nonnormal; (ii) any nonnormal components of the errors of measurements, $X_{ik} - \xi_{ik}$ and $Y_{ik} - \eta_{ik}$, are mutually independent, independent of ξ_{ik} and of the normal components of these errors; (iii) the normal components of the errors may be correlated but as a pair are independent of ξ_{ik}.

1. Introduction. Let ξ and η be two random variables known to be linearly connected, so that there exist two numbers, θ^* and p,

$$-\tfrac{1}{2}\pi < \theta^* \leqq \tfrac{1}{2}\pi, \qquad -\infty < p < +\infty, \tag{1}$$

such that the simultaneous values of ξ and η satisfy the condition

$$\xi \cos \theta^* + \eta \sin \theta^* - p = 0. \tag{2}$$

We consider the case where ξ and η are not directly observable but where the observations yield the simultaneous values of two other random variables X and Y, connected with ξ and η by the equations

$$X = \xi + U, \qquad Y = \eta + V. \tag{3}$$

Here U and V are unobservable random variables interpreted as errors in measuring ξ and η, respectively. Equation (2) is described as the linear structural relation between the variables X and Y. Throughout the paper it is assumed that the errors U and V may be correlated or not but, as a pair, are independent

[1] This paper was prepared with the partial support of the Office of Naval Research. It presents an extension of the contents of the Second Rietz Memorial Lecture delivered by the author at the Summer Meeting of the Institute of Mathematical Statistics at Boulder, Colorado, September 1st, 1949.

of the variables ξ and η. The problem considered is that of using a sequence $\{X_m, Y_m\}$ of completely independent pairs of observations on X and Y to construct a consistent estimate of θ^*. This is an old problem and a number of the earlier attempts to solve it are described by Wald in an important paper [1].

Early attempts to obtain a consistent estimate of θ^* were based exclusively on the sample variances and covariance of X and Y. However, as early as 1916, Godfrey Thomson showed [2] that the same first and second moments of the simultaneous distribution of X and Y are compatible with an infinity of different values of θ^* and that, therefore, attempts to estimate this parameter using only second order sample moments are doomed to failure. The writings of Thomson appear to have been overlooked and more and more studies were published using sample moments of the first and second orders as basic functions on which the estimates of θ^* were built. In 1936 [3] it was pointed out that, should it happen that the unobservable random variables ξ and η and also the errors U and V are normally distributed, then no consistent estimate of θ^* is possible because, in this event, the joint distribution of X and Y is also normal, and is determined by moments of the first two orders. Since these moments are consistent with an infinity of different values of θ^*, the latter is nonidentifiable. Between 1936 and the appearance of the paper by Wald in 1940 several studies were published, of which we will mention one by R. G. D. Allen [4], adding more precision to the facts just described.

Wald's paper brought a new idea into the situation. Namely, in certain cases something may be known about the particular values assumed by the unobservable random variable ξ. When this condition obtains, a method due to Wald gives a consistent estimate of θ^*. This estimate is again based on the arithmetic means of the observations on X and Y, appropriately grouped. Wald's idea took root and led to the paper by Housner and Brennan [5]. The same idea, a little more developed, is at the base of papers by Berkson [6] and by Hemelrijk [7]. However, important as these developments may be in various fields of application, it is obvious that they do not constitute a solution of the original problem of estimating θ^* when no knowledge of the particular values assumed by the unobservable random variables is postulated [8].

A new era in the study of the problem began following the result of Reiersøl [9][2] who proved that the case of nonidentifiability of θ^* noted in 1936 is an exception rather than a rule. This discovery stimulated the paper by Scott [10] giving a consistent estimate of θ^* applicable in a new category of cases, when no information on the particular values of ξ is postulated. However, the consistency of the estimate of Scott depends on the existence of a certain number of moments of the variable ξ.

The present paper is concerned with the case where the errors of measurement may be split into two components

[2] Although this paper appeared in print in 1950, the author became acquainted with it in the spring of 1948 from a lecture delivered by Reiersøl in a seminar meeting at the Statistical Laboratory, University of California, Berkeley.

$$
(4) \qquad \begin{cases} U = U_1 + U_2, \\ V = V_1 + V_2, \end{cases}
$$

where U_1 and V_1 are mutually independent and, as a pair, are independent of (U_2, V_2), and where U_2 and V_2 follow an arbitrary normal distribution. With the exception of the above independence, no restriction is placed on the distributions of U_1 and V_1. The purpose of the paper is to give an explicit construction of an estimate of a parameter θ (closely allied to but not identical with θ^*) which remains consistent in the most general case of identifiability, that is when ξ and η follow an arbitrary nonnormal distribution. No knowledge of particular values of ξ is postulated.

Since the above hypotheses admit the possibility that X and Y have no moments at all, the conventional methods of constructing the estimate have to be abandoned. Essentially, the estimate is defined as the abscissa which corresponds to the minimum ordinate of a point on a random curve. A search for this minimum among the roots of the derivative may be embarrassing. In fact, the derivative need not exist at all points. Therefore, the estimate is defined as the outcome of a specially devised interpolation procedure. The proof is based on a lemma which seems to have an interest of its own and may be applicable in other cases.

2. Concepts of identifiability and of consistent estimability. In order to define the concepts of identifiability[3] and of consistent estimability, we shall consider a variable point ϑ (parameter) capable of assuming any one of a set s of positions ϑ'. Every $\vartheta' \varepsilon s$ will be described as a possible value of ϑ. For every $\vartheta' \varepsilon s$ consider a specified set $\omega(\vartheta')$ of distribution functions and let ω stand for the union of all $\omega(\vartheta')$ for $\vartheta' \varepsilon s$.

DEFINITION 1. *We shall say that the parameter ϑ is identifiable in ω if, whatever $\vartheta' \varepsilon s$ and whatever $\vartheta'' \varepsilon s$, $\vartheta' \neq \vartheta''$, the corresponding sets $\omega(\vartheta')$ and $\omega(\vartheta'')$ have no elements in common.*

If ϑ is identifiable in ω, then to every distribution function $F \varepsilon \omega$ there corresponds a uniquely defined value of ϑ, say $\vartheta(F) \varepsilon s$.

From now on we shall restrict ourselves to sets ω of distribution functions F defined in the same Euclidean space of a fixed number m of dimensions. For every $F \varepsilon \omega$ we shall consider an m-dimensional random variable $X(F)$ whose distribution function is F. For $n = 1, 2, \cdots$ the symbol $Y_n(F)$ will denote the set of n completely independent observations made on $X(F)$. Thus, $Y_n(F)$ may be considered as a random variable of dimensionality mn. Let y_n denote a point in the mn-dimensioned Euclidean space R_{mn}. Consider a sequence of Borel measurable functions $\{T_n(y_n)\}$, each from R_{mn} to s. Obviously, the result $T_n(Y_n(F))$ of substituting $Y_n(F)$ for y_n in $T_n(y_n)$ is a random variable.

DEFINITION 2. *If the parameter ϑ is identifiable in ω and if, whatever be $F \varepsilon \omega$,*

[3] Important discussion of this concept, in a slightly different form, is due to Koopmans and Reiersøl [11]. This paper contains a substantial bibliography.

the sequence $\{T_n(Y_n(F))\}$ *converges in probability to* $\vartheta(F)$ *as* $n \to \infty$, *then this sequence is called a consistent estimate of* ϑ *in* ω.

DEFINITION 3. *If the parameter* ϑ *is identifiable in* ω *and if there exists a consistent estimate of* ϑ *in* ω, *then we shall say that* ϑ *is consistently estimable in* ω.

3. Identifiability of the directional parameter in the linear structural relation of two random variables. Returning to the general situation described in Section 1, denote by θ the parameter defined as follows:

$$\text{if} \quad -\tfrac{1}{2}\pi < \theta^* < \tfrac{1}{2}\pi, \quad \text{then} \quad \theta = \theta^*,$$

$$\text{if} \quad \theta^* = \tfrac{1}{2}\pi, \quad \text{then} \quad \theta = 0.$$

The parameter θ thus defined will be called the directional parameter of the structural relation (2).

Denote by S the set of possible values of θ, $-\frac{1}{2}\pi < \theta < \frac{1}{2}\pi$. For every value ϑ of this set we shall now define a set $\Omega(\vartheta)$ of joint distributions of the variables X and Y of formulae (3). We begin by defining $\Omega(0)$.

If $\theta = 0$ then either $\theta^* = 0$ or $\theta^* = \frac{1}{2}\pi$. Accordingly, $\Omega(0)$ is defined as the union of the two sets of distributions, $\Omega^*(0)$ and $\Omega^*(\frac{1}{2}\pi)$, each corresponding to a particular value of θ^*. If $\theta^* = 0$ then formula (2) implies that ξ is degenerate and $\xi = p$. Assume the following hypotheses:

(a) The variable η has an arbitrary distribution.

(b) $U = U_1 + U_2$ and $V = V_1 + V_2$, where U_1 and V_1 are mutually independent, as a pair are independent of ξ and η but otherwise arbitrarily distributed, and where (U_2, V_2) represent a pair of arbitrary normal variables, independent of the triplet η, U_1, V_1. In particular, U_2 and V_2 may be correlated.

(c) $-\infty < p < +\infty$.

Obviously, every specific set of hypotheses regarding η, U_1, V_1, U_2, V_2 and p implies a specific distribution of the pair X, Y. Then $\Omega^*(0)$ denotes the set of all such distributions.

In order to define $\Omega^*(\frac{1}{2}\pi)$ we notice that, if $\theta^* = \frac{1}{2}\pi$ then (2) implies that η is degenerate and $\eta = p$. $\Omega^*(\frac{1}{2}\pi)$ is defined to contain every joint distribution of X and Y implied by an arbitrary assumption regarding the distribution of ξ and by hypotheses (b) and (c).

As mentioned $\Omega(0)$ is the union of $\Omega^*(0)$ and $\Omega^*(\frac{1}{2}\pi)$. However, the reader will verify easily that the sets $\Omega^*(0)$ and $\Omega^*(\frac{1}{2}\pi)$ coincide. Therefore θ^* is not identifiable in $\Omega(0)$.

For every possible value ϑ of θ, other than $\vartheta = 0$, the set $\Omega(\vartheta)$ is defined to contain every joint distribution of X and Y defined by formulae (3), implied by the assumption that ξ follows an arbitrary nondegenerate, nonnormal distribution, that η is connected with ξ by equation (2) with p having an arbitrary real value, and that the errors U and V are arbitrarily distributed, subject to condition (b). It will be seen that the equality $\theta = 0$ characterizes the case where at least one of the variables ξ and η is degenerate so that, instead of being linearly connected, these variables may be considered as mutually independent.

Reiersøl proved [9] that the parameter θ is identifiable in the set Ω of distributions of X and Y defined as the union of all sets $\Omega(\vartheta)$ for $-\frac{1}{2}\pi < \vartheta < \frac{1}{2}\pi$. Since it is known that the restriction of nonnormality imposed on ξ and η when $\vartheta \neq 0$ cannot be relaxed without destroying the identifiability of θ, it follows that Ω is the broadest set of joint distributions of X and Y within which θ is identifiable, consistent with the assumption that the errors of measurement U and V satisfy assumption (b). The purpose of the present paper is to provide an explicit construction of an estimate of θ consistent in Ω.

4. A few preliminaries. It will be convenient to use the concept of uniform convergence in probability. Let $G(x)$ denote a function defined over a nondegenerate closed interval $x \varepsilon [a, b]$. Further, let $\{Z_n\}$ be a infinite sequence of random variables and $\{G_n(Z_n, x)\}$ a sequence of functions of two arguments Z_n and x. Each $G_n(Z_n, x)$ is assumed to be defined for every $x \varepsilon [a, b]$ and for every possible value of the random variable Z_n. Furthermore, when x is fixed, $G_n(Z_n, x)$ is a Borel measurable function of Z_n. Thus, it is a random variable.

DEFINITION 4. *We shall say that the sequence* $\{G_n(Z_n, x)\}$ *of random functions converges in probability to* $G(x)$ *uniformly in* $[a, b]$, *if there exists a function* $m(n)$ *defined for all* $n = 1, 2, \cdots$ *such that*

$$\lim_{n\to\infty} m(n) = \infty \tag{5}$$

and such that, whatever $\epsilon > 0$,

$$\lim_{n\to\infty} \left(m(n) \sup_{x\varepsilon[a,b]} P\{ \mid G_n(Z_n, x) - G(x) \mid > \epsilon \} \right) = 0. \tag{6}$$

Every function $m(n)$ satisfying the above conditions will be described as the norm of uniform convergence of $\{G_n(Z_n, x)\}$. Obviously, it may always be assumed that the norm $m(n)$ assumes only positive integer values.

In order to illustrate this concept, assume that for every $x \varepsilon [a, b]$ and for every $n = 1, 2, \cdots$ we have

$$E[G_n(Z_n, x)] = G(x) \tag{7}$$

and that the variance $\sigma_n^2(x)$ of $G_n(Z_n, x)$ is bounded by

$$\sigma_n^2(x) \leqq \frac{1}{n} \sigma_0^2, \tag{8}$$

where $\sigma_0 > 0$ is a constant. Using the inequality of Bienaymé-Tchebycheff we may write

$$P\{ \mid G_n(Z_n, x) - G(x) \mid > \epsilon \} < \frac{\sigma_n^2(x)}{\epsilon^2} \leqq \frac{\sigma_0^2}{n\epsilon^2} \tag{9}$$

for every $x \varepsilon [a, b]$. Thus

$$\sup_{x\varepsilon[a,b]} P\{ \mid G_n(Z_n, x) - G(x) \mid > \epsilon \} < \frac{\sigma_0^2}{n\epsilon^2}, \tag{10}$$

and it is seen that, under conditions (7) and (8), the sequence $\{G_n(Z_n, x)\}$ converges in probability to $G(x)$ uniformly in $[a, b]$. For example, the norm of uniform convergence may be defined as the greatest integer not exceeding the square root of n,

$$m(n) = [\sqrt{n}]. \tag{11}$$

Another convenient concept will be described as the m-lattice minimal point of a function f. This is defined as follows. Let $[a, b]$ denote a nondegenerate closed interval and $f(x)$ a real function defined on $x \varepsilon [a, b]$. Let m be an arbitrary integer $m > 1$ and

$$a_{mk} = a + k\frac{b-a}{m-1} \qquad \text{for} \quad k = 0, 1, \cdots, m-1. \tag{12}$$

We shall say that the m points a_{mk} form the m-lattice on $[a, b]$. Now consider the values $f(a_{mk})$ of $f(x)$ corresponding to the points of the m-lattice and use the symbol f_m to denote the smallest of these. In general, there will be r points of the lattice, say

$$a_{mk_1} < a_{mk_2} < \cdots < a_{mk_r} \tag{13}$$

such that $f(a_{mk_i}) = f_m$. Let $\mu = [(r+1)/2]$. The point $a_{m\mu}$ will be described as the m-lattice minimal point of the function $f(x)$. It will be denoted by $M_m(f(x))$.

Fundamental Lemma. *If the real function $G(x)$ is defined and continuous on a nondegenerate closed interval $[a, b]$ in which it has an absolute minimum $G(x_0)$ attained at a single point x_0, if $\{Z_n\}$ is a sequence of random variables and if the sequence of real random functions $\{G_n(Z_n, x)\}$ converges to $G(x)$ uniformly in $[a, b]$ with an integer valued norm $m(n)$, then the sequence $\{M_{m(n)}[G_n(Z_n, x)]\}$ of $m(n)$-lattice minimal points of $G_n(Z_n, x)$ converges in probability to x_0.*

Proof. Assume that the conditions of the lemma are satisfied. The proof consists in showing that, whatever $\epsilon > 0$ and $\eta > 0$, a number $N(\epsilon, \eta)$ can be found such that the inequality $n > N(\epsilon, \eta)$ implies

$$P\{| M_{m(n)}[G_n(Z_n, x)] - x_0 | > \epsilon\} < \eta. \tag{14}$$

Fix ϵ and η and denote by g the minimum value of $G(x)$ attained in the part of $[a, b]$ outside of the open interval $| x - x_0 | < \epsilon$. Obviously $g > G(x_0)$. Let $\delta < \epsilon$ be a sufficiently small positive number such that $| x - x_0 | < \delta$ implies

$$G(x_0) \leqq G(x) < G(x_0) + \tfrac{1}{3}(g - G(x_0)). \tag{15}$$

Denote by N_1 the smallest integer such that $n > N_1$ implies

$$\frac{b-a}{m(n)-1} < \delta \tag{16}$$

and by N_2 the smallest integer such that $n > N_2$ implies

$$m(n) \sup_{x \varepsilon [a,b]} P\{| G_n(Z_n, x) - G(x) | > \tfrac{1}{3}(g - G(x_0))\} < \eta. \tag{17}$$

Finally, let $N(\epsilon, \eta) = \max (N_1, N_2)$. It is easy to see that for $n > N(\epsilon, \eta)$, the inequality (14) is satisfied. We notice first that with $n > N(\epsilon, \eta) \geqq N_1$ the interval $(x_0 - \delta, x_0 + \delta)$ will include some points of the $m(n)$-lattice. Further, in order that $| M_{m(n)}[G_n(Z_n, x)] - x_0 | > \epsilon$ it is necessary that at least one of the values of $G_n(Z_n, x)$ assumed at points of the $m(n)$-lattice outside of the interval $(x_0 - \epsilon, x_0 + \epsilon)$ not exceeded any of the values assumed by this function on the $m(n)$-lattice within $(x_0 - \delta, x_0 + \delta)$. But outside of $(x_0 - \epsilon, x_0 + \epsilon)$ we have

$$G(x_0) < g \leqq G(x) \tag{18}$$

and inside of $(x_0 - \delta, x_0 + \delta)$

$$G(x) < G(x_0) + \tfrac{1}{2}(g - G(x_0)). \tag{19}$$

It follows that, if at each point of the $m(n)$-lattice the random function $G_n(Z_n, a_{mk})$ differs from $G(a_{mk})$ by at most $\frac{1}{2}(g - G(x_0))$, then

$$| M_{m(n)}[G_n(Z_n, x)] - x_0 | \leqq \epsilon.$$

Thus, the probability that $| M_{m(n)}[G_n(Z_n, x)] - x_0 | > \epsilon$ is at most equal to the probability, say π, that for at least one point a_{mk} of the $m(n)$-lattice $| G_n(Z_n, a_{mk}) - G(a_{mk}) | > \frac{1}{2}(g - G(x_0))$. However,

$$\begin{aligned} \pi &\leqq \sum_{k=0}^{m(n)-1} P\{| G_n(Z_n, a_{mk}) - G(a_{mk}) | > \tfrac{1}{2}(g - G(x_0))\} \\ &\leqq m(n) \sup_{x \varepsilon [a,b]} P\{| G_n(Z_n, x) - G(x) | > \tfrac{1}{2}(g - G(x_0))\} < \eta \end{aligned} \tag{20}$$

because of (17), and the proof of the lemma is completed.

5. Consistent estimates of the directional parameter of a linear structural relation between two variables. We return to the problem of the consistent estimation of the directional parameter θ of the structural relation (2). The parameter θ was defined in Section 3. Also it will be assumed that the joint distribution function F of the variables X and Y belongs to the set Ω defined in Section 3. Consider a set of $N(n) = 8n$ independent observations to be made on the pair of variables X and Y. These observations will be divided into n eight-tuples and denoted by (X_{ij}, Y_{ij}) for $i = 1, 2, \cdots, n$ and $j = 1, 2, \cdots, 8$. The ith eight-tuple will be denoted by Z_i^*. The totality of n eight-tuples will be denoted by Z_n.

In defining the estimate of θ we shall need three (identical or different) probability density functions $w_1(x)$, $w_2(x)$, and $w_3(x)$, and their characteristic functions, say $\Phi_1(t)$, $\Phi_2(t)$, and $\Phi_3(t)$, respectively. These probability density functions can be selected arbitrarily out of a class Γ which we shall define by the following conditions: every $w(x) \varepsilon \Gamma$ is symmetric about zero, $w(-x) = w(x)$, and there exists a positive number a such that $w(x) > 0$ for every $| x | < a$.

It will be observed that the symmetry of $w_k(x)$ implies that the corresponding characteristic function $\Phi_k(t)$ is real.

Speaking in terms of the characteristic functions Φ_1, Φ_2, Φ_3, we shall define a class C of consistent estimates of θ. Any particular choice of the functions Φ_1, Φ_2 and Φ_3 will determine a particular estimate of the class C. For example, we may choose to consider the following probability densities of class Γ: (1) the normal probability density with zero mean and unit variance, (2) the Cauchy probability density with unit scale and zero location parameter, and (3) the rectangular probability density between $-a$ and $+a$. Each of the corresponding characteristic functions, $\exp\{-\frac{1}{2}t^2\}$, $\exp\{-|t|\}$, and $\sin at/at$, respectively, may be taken to represent either Φ_1 or Φ_2 or Φ_3, or any two, or all three $\Phi_1 = \Phi_2 = \Phi_3$.

Assume that the choice of the functions $\Phi_k(t)$ is made. Denote by ϑ an arbitrary number between the limits $-\frac{1}{2}\pi \leqq \vartheta \leqq +\frac{1}{2}\pi$. For the kth eight-tuple of observations define the following symbols

$$\tag{21} \begin{cases} A(Z_k^*, \vartheta) = \Phi_1[(X_{k1} - X_{k2} + X_{k3} - X_{k4}) \cos \vartheta \\ \qquad + (Y_{k1} - Y_{k2} + Y_{k3} - Y_{k4}) \sin \vartheta]\Phi_2(X_{k1} - X_{k2} + X_{k5} - X_{k6}), \\ B(Z_k^*) = \Phi_3(Y_{k1} - Y_{k2} + Y_{k7} - X_{k8}), \\ C(Z_k^*) = \Phi_3(Y_{k1} - Y_{k4} - Y_{k6} + Y_{k7}), \\ D(Z_k^*) = \Phi_3(Y_{k3} - Y_{k4} + Y_{k5} - Y_{k6}), \end{cases}$$

$$\tag{22} H(Z_k^*, \vartheta) = A(Z_k^*, \vartheta)\{B(Z_k^*) - 2C(Z_k^*) + D(Z_k^*)\}.$$

Finally, let

$$\tag{23} G_n(Z_n, \vartheta) = \frac{1}{n} \sum_{k=1}^{n} H(Z_k^*, \vartheta).$$

Put $m(n) = [\sqrt{n}]$ and consider the $m(n)$-lattice on the closed interval $[-\frac{1}{2}\pi, +\frac{1}{2}\pi]$. For every fixed value Z_n' of Z_n we consider $G_n(Z_n', \vartheta)$ as a function of $\vartheta \in [-\frac{1}{2}\pi, +\frac{1}{2}\pi]$ and then $M_{m(n)}(G_n(Z_n', \vartheta))$ will denote its $m(n)$-lattice minimal point. After these preliminaries we define the estimate $T_n(Z_n)$ of θ as follows.

$$\tag{24} \begin{aligned} &\text{(i) If } G_n(Z_n, 0) \leqq \frac{1}{\sqrt[4]{n}}, \quad \text{then} \quad T_n(Z_n) = 0. \\ &\text{(ii) If } G_n(Z_n, 0) > \frac{1}{\sqrt[4]{n}}, \quad \text{then} \quad T_n(Z_n) = M_{m(n)}[G_n(Z_n, \vartheta)]. \end{aligned}$$

THEOREM. *The sequence $\{T_n(Z_n)\}$ represents an estimate of θ consistent in Ω.*

PROOF. We begin by noticing that, since the symbols in (21) are defined in terms of characteristic functions, their absolute values cannot exceed unity. Therefore,

$$\tag{25} |H(Z_k^*, \vartheta)| \leqq 4,$$

and thus all moments of $H(Z_k^*, \vartheta)$ exist. In particular, we shall be interested in the first moment, say

$$E\{H(Z_k^*, \vartheta)\} = E\{G_n(Z_n, \vartheta)\} = G(\vartheta), \tag{26}$$

and in the variance, say $\sigma^2(\vartheta)$, of $H(Z_k^*, \vartheta)$. Obviously, $\sigma^2(\vartheta) \leqq 16$. Since the successive variables $H(Z_k^*, \vartheta)$ are completely independent, the variance of $G_n(Z_n, \vartheta)$, say $\sigma_G^2(\vartheta)$, is

$$\sigma_G^2(\vartheta) = \frac{\sigma^2(\vartheta)}{n} \leqq \frac{16}{n}, \tag{27}$$

and it follows that the sequence $\{G_n(Z_n, \vartheta)\}$ converges in probability to $G(\vartheta)$ uniformly in $[-\frac{1}{2}\pi, \frac{1}{2}\pi]$. As we have seen before (see Section 4) the function $m(n) = [\sqrt{n}]$ may be taken as the norm of the uniform convergence.

Our next step in the proof consists in showing that the function $G(\vartheta)$ has the following properties.

(A) If the random variables X and Y follow a distribution $F \varepsilon \Omega$ such that $\theta(F) = 0$, then $G(\vartheta) \equiv 0$ for all $\vartheta \varepsilon [-\frac{1}{2}\pi, \frac{1}{2}\pi]$ including $\vartheta = 0$.

(B) If $\theta(F) \neq 0$, then $G(\vartheta) > 0$ for all $\vartheta \varepsilon [-\frac{1}{2}\pi, \frac{1}{2}\pi]$ with the exception of $\vartheta = \theta(F)$ where $G[\theta(F)] = 0$.

(C) $G(\vartheta)$ is continuous for $\vartheta \varepsilon [-\frac{1}{2}\pi, \frac{1}{2}\pi]$.

Once these three properties of $G(\vartheta)$ are established, the proof of the theorem is completed as follows. Assume first that $\theta(F) = 0$. Then, by the theorem of Bienaymé-Tchebycheff,

$$\begin{aligned} P\left\{G_n(Z_n, 0) \leqq \frac{1}{\sqrt[4]{n}}\right\} &\geqq P\left\{|G_n(Z_n, 0)| \leqq \frac{1}{\sqrt[4]{n}}\right\} \\ &> 1 - \sigma_G^2(0)\sqrt{n} \geqq 1 - \frac{16}{\sqrt{n}}. \end{aligned} \tag{28}$$

The definition of $T_n(Z_n)$ implies that it is equal to zero whenever

$$G_n(Z_n, 0) \leqq \frac{1}{\sqrt[4]{n}}, \quad \text{unconditionally,} \tag{29}$$

and also whenever

$$G_n(Z_n, 0) > \frac{1}{\sqrt[4]{n}} \quad \text{and} \quad M_{m(n)}[G_n(Z_n, \vartheta)] = 0. \tag{30}$$

Consequently, the probability

$$P\{T_n(Z_n) = 0\} \geqq P\left\{G_n(Z_n, 0) \leqq \frac{1}{\sqrt[4]{n}}\right\} \geqq 1 - \frac{16}{\sqrt{n}} \tag{31}$$

and tends to unity as $n \to \infty$.

Assume now that $\theta(F) \neq 0$. According to the fundamental lemma, in this case $M_{m(n)}(G_n(Z_n, \vartheta))$ converges in probability to $\theta(F)$. To prove that the same

is true for $T_n(Z_n)$ it is sufficient to show that the probability $P\{T_n(Z_n) \neq M_{m(n)}[G_n(Z_n, \vartheta)]\}$ tends to zero as $n \to \infty$. Obviously this last probability does not exceed the probability that $G_n(Z_n, 0) \leqq n^{-\frac{1}{4}}$. According to property (B) we have $G(0) > 0$ in the case considered. When $n > G(0)^{-4}$, we have

$$P\left\{G_n(Z_n, 0) \leqq \frac{1}{\sqrt[4]{n}}\right\} \leqq P\left\{\,|G_n(Z_n, 0) - G(0)| > G(0) - \frac{1}{\sqrt[4]{n}}\right\} < \frac{16}{n\left(G(0) - \frac{1}{\sqrt[4]{n}}\right)^2}, \tag{32}$$

and it follows that, as $n \to \infty$, the probability that $T_n(Z_n)$ will coincide with $M_{m(n)}(G_n(Z_n, \vartheta))$ tends to unity. It is seen that the properties (A), (B), and (C) of the function $G(\vartheta)$ combined with (26) imply that, whatever $F \varepsilon \Omega$, the estimate $\{T_n(Z_n)\}$ converges in probability to $\theta(F)$ or, in other words, that $T_n(Z_n)$ is an estimate of θ consistent in Ω. Therefore, in order to prove the theorem, we shall establish that the expectation (26) has the properties (A), (B), and (C). This will be done in Section 6 in the following order. First we shall use the postulated properties of the observable random variables X and Y and define a function $G(\vartheta)$ having the properties (A), (B), and (C). Next we shall show that the function $G(\vartheta)$ so defined coincides with the expectation (26).

6. Structural definition of $G(\vartheta)$. The structural definition of $G(\vartheta)$ is based on the properties of the characteristic function, say $\phi(t_1, t_2)$, of the joint distribution of X and Y. According to the usual definition

$$\phi(t_1, t_2) = E(e^{it_1X+it_2Y}), \tag{33}$$

where

$$\begin{aligned} X &= \xi + U_1 + U_2, \\ Y &= \eta + V_1 + V_2. \end{aligned} \tag{34}$$

Assume first that $\theta = 0$. In this case the components $\xi + U_1$ and $\eta + V_1$ are mutually independent and the possible dependence of X and Y will be due to the correlation that may exist between the normal components of errors U_2 and V_2. Since the logarithm of the characteristic function of two normal variables is a polynomial of the second order, when $\theta = 0$ the characteristic function of X and Y has the form, say

$$\phi(t_1, t_2 \mid \theta = 0) = e^{\psi_1(t_1)+\psi_2(t_2)+\gamma t_1 t_2}, \tag{35}$$

where $\psi_i(t_i)$ is a function of t_i alone, $i = 1, 2$. We note this form of $\phi(t_1, t_2 \mid \theta = 0)$ for future reference and proceed to the next case, where $\theta \neq 0$.

In this case $\theta = \theta^*$ and the structural relation (2) may be solved with respect to

$$\eta = \frac{p}{\sin \eta} - \xi \cot \theta. \tag{36}$$

Substituting this expression into (34) and denoting the logarithm of the characteristic function of ξ by $\chi(t)$, we have

$$\phi(t_1, t_2) = e^{\chi(t_1 - t_2 \cot \theta) + \psi_1(t_1) + \psi_2(t_2) + \gamma t_1 t_2}, \tag{37}$$

where the symbols ψ_1 and ψ_2 are again used to denote functions of one argument only, either t_1 or t_2. These functions in (37) have a meaning different from that in (35). However, this difference is of no importance because in both cases the essential point is that ψ_1 depends on t_1 but not on t_2 and that ψ_2 depends on t_2 but not on t_1. It will be convenient to consider that $\phi(t_1, t_2)$ always has the form (37) with the understanding that, when $\theta = 0$, then $\chi(t) \equiv 0$.

Since $\psi_1(t)$, $\psi_2(t)$ and $\chi(t)$ are defined in terms of logarithms of characteristic functions, they vanish at $t = 0$ and are continuous at this point. In addition, we shall use the following important property of $\chi(t)$. This is that, whenever $\theta(F) \neq 0$, then however small $\delta > 0$, the function $\chi(t)$ cannot coincide with a polynomial of second order on the whole of the interval $(-\delta, \delta)$. This property is implied by the hypothesis that, whenever $\theta \neq 0$ and therefore $\chi(t) \not\equiv 0$, then ξ is not normally distributed. In fact, assume that there exists a positive number δ^* such that $\chi(t) = a + bt + ct^2$ for all $|t| < \delta^*$. It is easy to see that in this case all the derivatives of the characteristic function of ξ would exist at $t = 0$ and would determine all the moments of ξ. Furthermore, these moments would coincide with the moments of a normal distribution, from which it would follow that ξ itself is normally distributed, contrary to the hypothesis. Thus it follows that, if $\chi(t)$ coincides with a quadratic in t over an interval, this interval cannot include $t = 0$.

Select a number $\vartheta \in [-\frac{1}{2}\pi, \frac{1}{2}\pi]$ and three arbitrary real numbers t, τ_1, τ_2. We shall consider $\phi(t_1, t_2)$ at the following eight points which, to abbreviate the formulae, will be denoted by lower case Roman numerals. Thus, for example, $\phi(\text{i})$ will denote the value of $\phi(t_1, t_2)$ evaluated at the first of the eight points. The coordinates of the first four points are

(i) $\quad t_1 = t \cos \vartheta + \tau_1, \quad t_2 = t \sin \vartheta + \tau_2,$

(ii) $\quad t_1 = t \cos \vartheta + \tau_1, \quad t_2 = t \sin \vartheta,$

(iii) $\quad t_1 = t \cos \vartheta, \quad t_2 = t \sin \vartheta + \tau_2,$

(iv) $\quad t_1 = t \cos \vartheta, \quad t_2 = t \sin \vartheta.$

The coordinates of points (v) through (viii) are obtained from those of (i) to (iv), respectively, by substituting $t = 0$. Thus

(v) $\quad t_1 = \tau_1, \quad t_2 = \tau_2,$

(vi) $\quad t_1 = \tau_1, \quad t_2 = 0,$

(vii) $\quad t_1 = 0, \quad t_2 = \tau_2,$

(viii) $\quad t_1 = 0, \quad t_2 = 0.$

Obviously $\phi(\text{viii}) = 1$. Now we form the function

$$h(\vartheta, t, \tau_1, \tau_2) = \phi(\text{i})\phi(\text{iv})\phi(\text{vi})\phi(\text{vii}) - \phi(\text{ii})\phi(\text{iii})\phi(\text{v})\phi(\text{viii}). \tag{38}$$

Easy algebra gives

$$h(\vartheta, t, \tau_1, \tau_2) = \Psi_1\Psi_2 - \Psi_1\Psi_3, \tag{39}$$

where

$$\begin{cases} \Psi_1 = \exp\{\psi_1(t\cos\vartheta + \tau_1) + \psi_1(t\cos\vartheta) + \psi_1(\tau_1) \\ \qquad\qquad + \psi_2(t\sin\vartheta + \tau_2) + \psi_2(t\sin\vartheta) + \psi_2(\tau_2) \\ \qquad\qquad + \gamma[(t\cos\vartheta + \tau_1)(t\sin\vartheta + \tau_2) + t^2\cos\vartheta\sin\vartheta]\}, \\ \Psi_2 = \exp\{\chi(At + \tau_1 - \tau_2\cot\theta) + \chi(At) + \chi(\tau_1) + \chi(-\tau_2\cot\theta)\}, \\ \Psi_3 = \exp\{\chi(At + \tau_1) + \chi(0) + \chi(At - \tau_2\cot\theta) + \chi(\tau_1 - \tau_2\cot\theta)\}, \end{cases} \tag{40}$$

with

$$A = \frac{\sin(\theta - \vartheta)}{\sin\theta}. \tag{41}$$

For any $x > 0$ we shall use the symbol $\sigma(x)$ to denote the set of triplets (t, τ_1, τ_2) such that $|t| < x$, $|\tau_1| < x$ and $|\tau_2| < x$. Because of the properties of the functions ψ_1, ψ_2, and χ there exists a positive number δ such that within $\sigma(\delta)$ the functions Ψ_1 and Ψ_3 do not vanish. Consequently, for $(t, \tau_1, \tau_2) \varepsilon \sigma(\delta)$ we may write

$$\begin{aligned} h(\vartheta, t, \tau_1, \tau_2) &= \Psi_1\Psi_3\left(\frac{\Psi_2}{\Psi_3} - 1\right) \\ &= \Psi_1\Psi_3\Big(\exp\{[\chi(At + \tau_1 - \tau_2\cot\theta) - \chi(At + \tau_1) \\ &\qquad\qquad - \chi(At - \tau_2\cot\theta) + \chi(At)] \\ &\qquad - [\chi(\tau_1 - \tau_2\cot\theta) - \chi(\tau_1) - \chi(-\tau_2\cot\theta) + \chi(0)]\} - 1\Big). \end{aligned} \tag{42}$$

The idea of the function $h(\vartheta, t, \tau_1, \tau_2)$ originated from the paper of Reiersøl and this function is the key to the whole construction of the estimate $T_n(Z_n)$. The function $h(\vartheta, t, \tau_1, \tau_2)$ is defined as a combination of values of the characteristic function of the observable random variables X and Y at eight arbitrarily selected points. Consequently, the definition of $h(\vartheta, t, \tau_1, \tau_2)$ is independent of the value of $\theta(F)$. However, the properties of $h(\vartheta, t, \tau_1, \tau_2)$ do depend on $\theta(F)$, as follows.

(a) If $\theta(F) = 0$, then $h(\vartheta, t, \tau_1, \tau_2) = 0$ for all values of the four arguments $\vartheta \varepsilon [-\frac{1}{2}\pi, \frac{1}{2}\pi]$ and $-\infty < t, \tau_1, \tau_2 < +\infty$.

(b) If $\theta(F) \neq 0$ and $\vartheta = \theta(F)$, then $h(\vartheta, t, \tau_1, \tau_2) = 0$ for all combinations of values of t, τ_1, τ_2, $-\infty < t, \tau_1, \tau_2 < +\infty$.

(c) If $\theta(F) \neq 0$ and $\vartheta \neq \theta(F)$ then, whatever $\delta_1 > 0$, the cube $\sigma(\delta_1)$ contains a subset of points (t, τ_1, τ_2) of positive three-dimensional measure within which $h(\vartheta, t, \tau_1, \tau_2) \neq 0$.

In order to prove (a) we notice that the case $\theta(F) = 0$ is characterized by the identity $\chi(t) \equiv 0$. Making this substitution in (42) it is immediately seen that, in this particular case, $h(\vartheta, t, \tau_1, \tau_2) \equiv 0$ for all combinations of values of the four arguments.

In order to prove (b) we notice that $\vartheta = \theta(F)$ implies

$$A = \frac{\sin(\vartheta - \theta)}{\sin \theta} = 0. \tag{43}$$

Then (42) implies that $h(\theta, t, \tau_1, \tau_2) \equiv 0$ for all combinations of values of the three arguments t, τ_1, τ_2.

In proving (c) we shall use the hypothesis that ξ is not a normal variable and that, therefore, however small $\delta > 0$, the function $\chi(t)$ cannot coincide with a polynomial of second order on the whole of the interval $|t| < \delta$. Assume that the assertion (c) is not true and that, with $\vartheta \neq \theta(F) \neq 0$, there exists a positive number δ^* such that, for $(t, \tau_1, \tau_2) \in \sigma(\delta^*)$ we have identically $h(\vartheta, t, \tau_1, \tau_2) \equiv 0$. Then this identity will also hold for all sufficiently small $|t|$ and $|\tau_1|$ and

$$\tau_2 = -\tau_1 \tan \theta. \tag{44}$$

Within the common part of $\sigma(\delta)$ and $\sigma(\delta^*)$ the functions Ψ_1 and Ψ_3 do not vanish. Therefore, we must conclude that the result of substituting (44) into Ψ_2 and Ψ_3 must give $\Psi_2/\Psi_3 \equiv 1$ for all sufficiently small $|t|$ and $|\tau_1|$. This however, implies that

$$\chi(At + 2\tau_1) - 2\chi(At + \tau_1) + \chi(At) \equiv \chi(2\tau_1) - 2\chi(\tau_1) + \chi(0). \tag{45}$$

It will be seen that the expressions on both sides of this identity represent second differences of $\chi(t)$ at steps τ_1 evaluated at points At and zero, respectively. Thus, the assumption $h(\vartheta, t, \tau_1, \tau_2) \equiv 0$ in $(t, \tau_1, \tau_2) \in \sigma(\delta^*)$ leads to the conclusion that there must exist a certain vicinity W of the point $t = 0$ where, however small $|\tau_1|$, the second difference of the function $\chi(t)$ computed at steps τ_1 has a value possibly depending on τ_1 but not on the point at which it is evaluated. Since $\chi(t)$ is continuous, it must then coincide with a polynomial of second order in t over the whole interval W. This, however, is contrary to the hypothesis. Therefore, if $\vartheta \neq \theta(F) \neq 0$, whatever be $\delta > 0$ the cube $\sigma(\delta)$ must contain at least one point t', τ_1', τ_2' such that $h(\vartheta, t', \tau_1', \tau_2') \neq 0$. Since h is continuous in (t, τ_1, τ_2) it then follows that $\sigma(\delta)$ must contain a set of three-dimensional positive measure where $h(\vartheta, t, \tau_1, \tau_2) \neq 0$. This establishes (c).

When $h(\vartheta, t, \tau_1, \tau_2) \neq 0$, it may be represented by a real or by a complex number. It is known that by changing the signs of the arguments of any characteristic function one obtains a value which is conjugate to the original value of this characteristic function. It is easy to see that the same applies to $h(\vartheta, t, \tau_1, \tau_2)$. Therefore, the product

$$h(\vartheta, t, \tau_1, \tau_2)h(\vartheta, -t, -\tau_1, -\tau_2) = |h(\vartheta, t, \tau_1, \tau_2)|^2 = g(\vartheta, t, \tau_1, \tau_2), \tag{46}$$

say, is equal to the square of the modulus of $h(\vartheta, t, \tau_1, \tau_2)$. It follows from the preceding that the function $g(\vartheta, t, \tau_1, \tau_2)$ is real valued, nonnegative and continuous in (t, τ_1, τ_2). Also, it is easy to see that $g(\vartheta, t, \tau_1, \tau_2)$ cannot be greater than 4. Furthermore, if $\theta(F) = 0$ then g is identically zero. Also, it is zero identically in t, τ_1, τ_2 if $\theta(F) \neq 0$ but $\vartheta = \theta(F)$.

On the other hand, if $\theta(F) \neq 0$ and $\vartheta \neq \theta(F)$, then in every vicinity of $t = \tau_1 = \tau_2 = 0$ there is a set of positive three-dimensional measure where $g(\vartheta, t, \tau_1, \tau_2) > 0$. Now, let $w_1(x)$, $w_2(x)$ and $w_3(x)$ be three (identical or different) probability density functions of class Γ (that is, each symmetric about $x = 0$ and nonvanishing in a nondegenerate interval $|x| < a$). Also, let

$$G(\vartheta) = \int_{-\infty}^{+\infty} w_1(t) \int_{-\infty}^{+\infty} w_2(\tau_1) \int_{-\infty}^{+\infty} w_3(\tau_2) g(\vartheta, t, \tau_1, \tau_2)\, dt\, d\tau_1\, d\tau_2. \tag{47}$$

It is obvious that, whatever the chosen probability density functions w_1, w_2, and w_3,

$$\text{if } \theta(F) = 0, \text{ then } G(\vartheta) = 0 \text{ for every } \vartheta \,\varepsilon\, [-\tfrac{1}{2}\pi, \tfrac{1}{2}\pi],$$

$$\text{if } \theta(F) \neq 0 \text{ and } \vartheta = \theta(F), \text{ then } G(\vartheta) = 0,$$

$$\text{if } \theta(F) \neq 0 \text{ and } \vartheta \neq \theta(F), \text{ then } G(\vartheta) > 0.$$

Also, because of the definition of $g(\vartheta, t, \tau_1, \tau_2)$ in terms of the characteristic function of X and Y, $G(\vartheta)$ is a continuous function of ϑ. It follows that the function G defined in formula (47) possesses the properties (A), (B), and (C) mentioned at an earlier stage of the proof of the theorem (Section 5). In order to complete this proof, we now show that, if $\Phi_k(t)$ denotes the characteristic function of $w_k(x)$, $k = 1, 2, 3$, then the expectation of $H(Z_m^*, \vartheta)$, defined by (22) and (21), is equal to $G(\vartheta)$ of formula (47), identically in $\vartheta \,\varepsilon\, [-\tfrac{1}{2}\pi, \tfrac{1}{2}\pi]$.

For this purpose we return to the function $g(\vartheta, t, \tau_1, \tau_2)$ and reexamine its definition (46) in terms of $h(\vartheta, t, \tau_1, \tau_2)$ and ultimately in terms of the characteristic function $\phi(t_1, t_2)$ as in (38). It is seen that $g(\vartheta, t, \tau_1, \tau_2)$ and $G(\vartheta)$ may be written conveniently as linear combinations of four terms each, say

$$G(\vartheta) = G_1(\vartheta) - G_2(\vartheta) - G_3(\vartheta) + G_4(\vartheta), \tag{48}$$

$$g(\vartheta, t, \tau_1, \tau_2) = g_1 - g_2 - g_3 + g_4, \tag{49}$$

where, for $k = 1, 2, 3, 4$,

$$G_k(\vartheta) = \iiint g_k w_1(t) w_2(\tau_1) w_3(\tau_2)\, dt\, d\tau_1\, d\tau_2, \tag{50}$$

and g_k stands for the product of from six to eight factors, each factor representing the characteristic function of X and Y evaluated at specified values of the two arguments. Upon inspecting (46) and (38) the reader will have no difficulty in writing down the expressions of the four components g_k. To save space we shall

reproduce only the expression of g_1 represented by the product of eight factors, as follows:

$$
\begin{aligned}
g_1 = {} & \phi(t \cos \vartheta + \tau_1, t \sin \vartheta + \tau_2)\phi(-t \cos \vartheta - \tau_1, -t \sin \vartheta - \tau_2) \\
& \cdot \phi(t \cos \vartheta, t \sin \vartheta)\phi(-t \cos \vartheta, -t \sin \vartheta) \\
& \cdot \phi(\tau_1, 0)\phi(-\tau_1, 0)\phi(0, \tau_2)\phi(0, -\tau_2).
\end{aligned} \tag{51}
$$

Consider the kth eight-tuple of independent observations on X and Y and let (X_{kj}, Y_{kj}) represent the jth pair of this eight-tuple. Obviously, we may write

$$
\begin{aligned}
& \phi(t \cos \vartheta + \tau_1, t \sin \vartheta + \tau_2) \\
& \qquad = E[\exp \{it(X_{k1} \cos \vartheta + Y_{k1} \sin \vartheta) + i\tau_1 X_{k1} + i\tau_2 Y_{k1}\}],
\end{aligned} \tag{52}
$$

$$
\begin{aligned}
& \phi(-t \cos \vartheta - \tau_1, -t \sin \vartheta - \tau_2) \\
& \qquad = E[\exp \{-it(X_{k2} \cos \vartheta + Y_{k2} \sin \vartheta) - i\tau_1 X_{k2} - i\tau_2 Y_{k2}\}],
\end{aligned} \tag{53}
$$

etc. Because of the complete independence of all the eight pairs (X_{kj}, Y_{kj}), the expression of g_1 may be written as the expectation of a single exponential,

$$
\begin{aligned}
g_1 = {} & E[\exp \{it((X_{k1} - X_{k2} + X_{k3} - X_{k4}) \cos \vartheta \\
& + (Y_{k1} - Y_{k2} + Y_{k3} - Y_{k4}) \sin \vartheta) + i\tau_1(X_{k1} - X_{k2} + X_{k5} - X_{k6}) \\
& + i\tau_2(Y_{k1} - Y_{k2} + Y_{k7} - Y_{k8})\}].
\end{aligned} \tag{54}
$$

This expectation is just a convenient symbol for an eightfold Stieltjes integral with respect to the distribution function $F(x_j, y_j)$ of each pair (X_{kj}, Y_{kj}). Thus the component $G_1(\vartheta)$ of $G(\vartheta)$ is an elevenfold integral. Since this integral is absolutely convergent, we may invert the order of integration and write

$$
\begin{aligned}
G_1(\vartheta) = E\Bigg(& \int_{-\infty}^{+\infty} e^{it((X_1 - X_2 + X_3 - X_4)\cos\vartheta + (Y_1 - Y_2 + Y_3 - Y_4)\sin\vartheta)} w_1(t)\, dt \\
& \cdot \int_{-\infty}^{+\infty} e^{i\tau_1(X_1 - X_2 + X_5 - X_6)} w_2(\tau_1)\, d\tau_1 \\
& \cdot \int_{-\infty}^{+\infty} e^{i\tau_2(Y_1 - Y_2 + Y_7 - Y_8)} w_3(\tau_2)\, d\tau_2 \Bigg),
\end{aligned} \tag{55}
$$

or, remembering the definition of Φ_1, Φ_2, and Φ_3,

$$
\begin{aligned}
G_1(\vartheta) = {} & E\{\Phi_1[(X_1 - X_2 + X_3 - X_4) \cos \vartheta \\
& + (Y_1 - Y_2 + Y_3 - Y_4) \sin \vartheta] \\
& \cdot \Phi_2(X_1 - X_2 + X_5 - X_6)\Phi_3(Y_1 - Y_2 + Y_7 - Y_8)\},
\end{aligned} \tag{56}
$$

or, finally

$$
G_1(\vartheta) = E[A(Z_k^*, \vartheta)B(Z_k^*)], \tag{57}
$$

with the symbols $A(Z_k^*, \vartheta)$ and $B(Z_k^*)$ defined for every eight-tuple of completely independent observations as in formulae (21). Similarly it is easy to show that

$$\begin{cases} G_2(\vartheta) = G_3(\vartheta) = E[A(Z_k^*, \vartheta)C(Z_k^*)], \\ G_4(\vartheta) = E[A(Z_k^*, \vartheta)D(Z_k^*)]. \end{cases} \tag{58}$$

This, however, implies that

$$G(\vartheta) = E[H(Z_k^*, \vartheta)], \tag{59}$$

and the proof of the theorem is completed.

7. Acknowledgment. The results presented in this paper differ in several respects from the contents of the Second Rietz Memorial Lecture of 1949. Among other things it was possible to remove a certain restrictiveness of the original estimate of θ. The parameter considered in 1949 was not θ itself but rather $\beta = \cot \theta$. In order to construct the original estimate of β, it was necessary to use a number B known to exceed $|\beta|$. It is a pleasure to acknowledge the author's indebtedness to Professor Charles M. Stein for a useful suggestion which led to the present construction of the estimate of θ, independent of any advance knowledge of the value of this parameter.

REFERENCES

[1] ABRAHAM WALD, "The fitting of straight lines if both variables are subject to error," *Annals of Math. Stat.*, Vol. 11 (1940), pp. 284–300.

[2] GODFREY H. THOMSON, "A hierarchy without a general factor," *British Jour. Psych.*, Vol. 8 (1916), pp. 271–281.

[3] J. NEYMAN, "Remarks on a paper by E. C. Rhodes," *Jour. Roy. Stat. Soc.*, Vol. 100 (1937), pp. 50–57.

[4] R. G. D. ALLEN, "The assumptions of linear regression," *Economica*, N. S., Vol. 6 (1939), pp. 191–204.

[5] W. G. HOUSNER AND J. F. BRENNAN, "The estimation of linear trends," *Annals of Math. Stat.*, Vol. 19 (1948), pp. 380–388.

[6] JOSEPH BERKSON, "Are there two regressions?" *Jour. Am. Stat. Assn.*, Vol. 45 (1950), pp. 164–180.

[7] J. HEMELRIJK, "Construction of a confidence region for a line," *Nederl. Akad. Wetensch., Proc.*, Vol. 52 (1949), pp. 995–1005.

[8] J. NEYMAN AND ELIZABETH L. SCOTT, "On certain methods of estimating the linear structural relation," *Annals of Math. Stat.*, Vol. 22 (1951), pp. 352–361.

[9] OLAV REIERSØL, "Identifiability of a linear relation between variables which are subject to error," *Econometrica*, Vol. 18 (1950), pp. 375–389.

[10] ELIZABETH L. SCOTT, "Note on consistent estimates of the linear structural relation between two variables," *Annals of Math. Stat.*, Vol. 21 (1950), pp. 284–288.

[11] T. C. KOOPMANS AND O. REIERSØL, "The identification of structural characteristics," *Annals of Math. Stat.*, Vol. 21 (1950), pp. 165–181.

TEST CRITERIA FOR HYPOTHESES OF SYMMETRY OF A REGRESSION MATRIX[1]

By Uttam Chand[2]

University of North Carolina and Boston University

Summary. Hotelling's [1] theoretical findings in mathematical economics on the rational behavior of buyers in maximizing their net profit indicate that the matrix of the first partial derivatives of a set of related demand functions would be symmetric and negative definite. It is the object of this paper to determine whether the assumption of symmetry will be tenable in the light of the particular set of observations. The study of test functions for the property of definiteness as a whole will form the subject of a forthcoming paper. The present investigation assumes that the demand functions are regression functions and, therefore, results obtained in this paper do not cover all types of demand functions. The test function U proposed here for the hypothesis of symmetry is invariant under all contragredient transformations. The distribution of U depends on unknown nuisance parameters. The likelihood ratio under the hypothesis of symmetry leads to a multilateral matric equation which represents $\frac{1}{2}\, p(p+1)$ equations of the third degree in $\frac{1}{2}\, p(p+1)$ unknown regression coefficients for the p-variate case. It has not been possible to establish the existence of a nontrivial solution of this equation, and it is, therefore, not being given here.

1. Introduction. Let p_i denote the price of the ith commodity and q_i the quantity consumed at that price. Consider $p_i = f_i(q_1, q_2, \cdots)$ a set of demand functions and let $u = u(q_1, q_2, \cdots)$ represent the gross receipts of a purchaser of goods. Under the assumption that each entrepreneur tries to maximize his net profit $\pi = u - \Sigma p_i q_i$, Hotelling [1] in an important contribution concerning the theoretical nature of supply and demand functions showed that if the entrepreneur is working in a steady economic state in which there is no restriction on his money expenditure, then the matrix of the first partial derivatives of prices on quantities would be symmetric, that is,

$$\frac{\partial p_i}{\partial q_j} = \frac{\partial p_j}{\partial q_i}$$

and that for a true maximum such a matrix would be negative definite, that is,

$$\frac{\partial p_i}{\partial q_i} < 0, \qquad \frac{\partial(p_i, p_j)}{\partial(q_i, q_j)} > 0, \qquad \frac{\partial(p_i, p_j, p_k)}{\partial(q_i, q_j, q_k)} < 0, \cdots.$$

[1] This paper was presented at the Cleveland meeting of the Institute on December 27, 1948.

[2] The author wishes to express his grateful appreciation to Professors Harold Hotelling and William G. Madow for guidance in this research.

It would thus appear that the inequalities arising out of the negative definiteness of the matrix generalize the conditions that a demand curve shall decline.

No suitable statistical tests have existed for testing the hypothesis of symmetry and negative definiteness of the matrix referred to in the previous paragraph. Henry Schultz [2] was first to consider such a question and the present paper has grown out of his statistical attempts. To verify Hotelling's laws on the basis of a particular set of data consider

$$p_i = f_i(q_1, q_2, \cdots) + u_i, \tag{1.1}$$

a system of demand equations where p_i and q_j denote current prices and quantities and where u_i is a stochastic variable. We shall assume that the quantities are fixed and prices are determined by demand. For example, some government agency could conduct actual experiments fixing alternative sets of quantities and observing what prices the choice of buyers would lead to. In such a situation we shall, therefore, be justified in assuming demand functions to be regression functions. In general the quantities are determined by a certain type of supply function under the prevailing market mechanism. Suppose the supply functions are given by

$$p_i = Q_i(q_1, q_2, \cdots) + v_i, \tag{1.2}$$

where v_i is a stochastic variable and u's and v's have a more or less specified joint probability law. If (1.1) and (1.2) are to hold simultaneously their solutions, if they exist, will be the only observable values of prices and quantities; and, therefore, quantities such as $\partial q_i/\partial p_j$ cannot in general be estimated and consequently no question of testing symmetry could be raised. However we could conceive of a different type of supply functions from those in (1.2) containing other independently determined variables besides the p's and being of such a stochastic type that the equations (1.1) would be regression equations [12]. For the purpose of this investigation we shall assume that the demand equations are regression equations such that the mathematical expectation of p_i is equal to f_i and since not all demand functions are regression equations, the results of the present investigation are not applicable to all types of demand equations.

Since we are studying certain properties of correlated variables any proposed statistical criterion must satisfy the property of invariance under linear transformations of prices and quantities. The fact of the transformation of quantities being not independent of that of prices will further restrict us to the consideration of such relations as are invariant under a linear transformation of one set of variates contragredient to those of the other ([3], pp. 108–109). The importance of such a class of relations was first suggested by Hotelling in a series of papers [4], [5], [6]. Examples of such a "value preserving" class of transformations may be found in the mixing of different grades of wheat or the combination of raw materials and labor into finished products such that the total value remains unchanged.

The statistic U (Section 4) proposed here for the hypothesis of symmetry for

the case of two related commodities is invariant under all contragredient transformations. It is exact in the sense that its probability distribution law is precisely determined under the hypothesis. Certain practically useful relations between this statistic and Student's t will be indicated. This test has in addition the property of being an unbiased test in the sense of Neyman and Pearson. We consider its p-variate generalization in Section 4.4.

2. Probability model. Let $Y = \| y_{i\alpha} \|$ be a $p \times N$ sample matrix from a normal multivariate parent having $\sigma = \| \sigma_{ij} \|$ as the dispersion matrix and $\eta = \| \eta_{i\alpha} \| = \beta X$ as the corresponding matrix of expectations where $\beta = \| \beta_{ij} \|$ is the population regression matrix and $X = \| x_{i\alpha} \|$ is the matrix of nonrandom observations on the fixed variates (e.g., $y_1, \cdots, y_p$ may denote prices and $x_1, \cdots, x_p$ the quantities consumed at these prices). Let $g = \| g_{ij} \| = XY'$, where $g_{ij} = \Sigma_\alpha x_{i\alpha} y_{j\alpha}$ and where Y' is the transpose of Y. Set $a = XX'$ and $c = \| c_{ij} \| = a^{-1}$ $(i, j = 1, \cdots, p; \alpha = 1, \cdots, N; p \leq N)$. We shall assume without loss of generality that $y_{i\alpha}$'s and $x_{i\alpha}$'s are either measured from their means or from polynomial means if the y's are subject to a time trend. It will be noticed that the symmetry and definiteness of the matrix of partial derivatives of prices on quantities is equivalent to the symmetry and definiteness of the regression matrix β.

3. Contragredient transformation of the two sets of variates. Let $f = \| f_{ij} \|$ be a $p \times p$ nonsingular matrix and let the columns x of the matrix X be subjected to the transformation f; we write $x^* = fx$. If the columns y are transformed into columns y^* in such a way that $y^{*\prime}x^* = y'x$ for every x and y, then the transformation of the y's is uniquely determined, viz., $y^* = f'^{-1}y$. We say, under these circumstances, that the columns x on the one hand, and the columns y on the other hand, are transformed contragrediently under f. For the mathematical expectation of y^*'s we have $E(y^*) = \beta^* x^*$ where $\beta^* = f'^{-1}\beta f^{-1}$. Consequently $\beta^{*\prime} = \beta^*$ implies $\beta' = \beta$ and conversely. Thus we notice that the symmetry of the matrix β is preserved by this type of transformation. Since the property of definiteness is invariant under any nonsingular linear transformation, the hypotheses of symmetry and definiteness are invariant and we might as well consider the properties of the matrix β^*. If we denote by $\sigma^* = \| \sigma^*_{ij} \|$ the covariance matrix of the y^*'s, we have $\sigma^* = f'^{-1}\sigma f^{-1}$ and consequently the ratio of the determinants $|\beta|$ and $|\sigma|$ is an absolute invariant. We now state the following theorem:

THEOREM I. *If σ is a positive definite matrix and β a real symmetric matrix and the two are cogrediently transformed, there exists a nonsingular linear transformation which will reduce σ to an identity matrix and β to a diagonal matrix.*

PROOF. We have $\beta = f'\beta^* f$ and $\sigma = f'\sigma^* f$ and the proof follows from a theorem given in [3] (p. 171). We shall make use of this result in Sections 4 and 5.

4. Hypothesis of symmetry of the regression matrix β.

4.1. *The statistic U.* We shall show that for the bivariate case the statistic U

now to be presently defined provides an exact and unbiased test for $H_0 : \beta_{12} = \beta_{21}$ against the set of alternatives which do not specify anything except $\beta_{12} \neq \beta_{21}$. Consider

$$U = (b_{12} - b_{21})^2(c_{22}s_{11} + c_{11}s_{22} - 2c_{12}s_{12})^{-1},$$

where

(i) The sample regression coefficients b_{ij} are normally distributed with means β_{ij} and $E(b_{kj} - \beta_{kj})(b_{mi} - \beta_{mi}) = \sigma_{km}c_{ij}$.

(ii) The s_{ij}'s are the unbiased estimates, each based on (say) n degrees of freedom, of σ_{ij} and follow the Wishart [7] law. Actually we have

$$ns_{ij} = \sum_{\alpha=1}^{N} (y_{i\alpha} - Y_{i\alpha})(y_{j\alpha} - Y_{j\alpha}),$$

where $Y_{i\alpha}$'s are sample regression functions.

(iii) The c_{ij}'s have been previously defined (Section 2).

Under the assumption of the conditional bivariate normal law for the $y_{i\alpha}$'s (Section 2), the residuals of $y_{i\alpha}$'s from their respective sample regression functions are normally distributed. If the $y_{i\alpha}$'s are subject to a time trend, as may very often be the case in economics, it will be more appropriate to consider the model

$$E(y_{i\alpha}) = \alpha_0 + \alpha_1\xi_1(t) + \alpha_2\xi_2(t) + \cdots + \beta_{i1}(x_{1\alpha} - \bar{x}_1) + \beta_{i2}(x_{2\alpha} - \bar{x}_2),$$

where the $\xi(t)$'s are known polynomials in time. Under such a model also residuals are known to be normally distributed. Consequently we might as well have assumed such a model which will thus only affect the number of degrees of freedom available for the estimates s_{ij}.

THEOREM II. *If x and y are transformed contragrediently, the statistic U is an absolute invariant.*

PROOF. Set $s = \| s_{ij} \|$ and $b = \| b_{ij} \|$. Under the contragredient transformation of x and y (Section 3) we have $b = f'b^*f$; $s = f's^*f$; and $c = f'c^*f$. If we perform this transformation on U and simplify, we notice that the numerator and denominator of U are relative invariants of weight -2 and consequently U is an absolute invariant.

4.2. *Distribution of U under the null hypothesis.* Since U is an absolute invariant under the contragredient transformation of x and y we may derive the distribution of U taking σ to be an identity matrix and β to be a diagonal matrix (Theorem I) in the parametric space.

The numerator and denominator of U are distributed independently of one another [8]. Let $Z = c_{22}s_{11} + c_{11}s_{22} - 2c_{12}s_{12}$. This is a positive definite quadratic form in normally distributed variates. Let u_α and v_α represent residuals of y_1 and y_2 from the corresponding sample regression functions. There exists an orthogonal transformation of the N variables u_α and v_α which will simultaneously yield $s_{11} = \Sigma_1^n u_\alpha^{*2}/n$; $s_{22} = \Sigma_1^n v_\alpha^{*2}/n$ and $s_{12} = \Sigma_1^n u_\alpha^* v_\alpha^*/n$, where u^* and v^* are normally

and independently distributed with 0 means and a common variance for each set u^*, v^*. Consider now the orthogonal transformation

$$u'_\alpha = u^*_\alpha \cos\theta - v^*_\alpha \sin\theta,$$

$$v'_\alpha = u^*_\alpha \sin\theta + v^*_\alpha \cos\theta,$$

where θ is so determined that $nZ = d_1\Sigma_1^n u'^2_\alpha + d_2\Sigma_1^n v'^2_\alpha$; then

$$d_1 = \tfrac{1}{2}\{c_{11} + c_{22} + [(c_{11} - c_{22})^2 + 4c_{12}^2]^{\frac{1}{2}}\},$$

$$d_2 = \tfrac{1}{2}\{c_{11} + c_{22} - [(c_{11} - c_{22})^2 + 4c_{12}^2]^{\frac{1}{2}}\}.$$

Consequently $U = \chi^2_{1,1}(C_1\chi^2_{2,n} + C_2\chi^2_{3,n})^{-1}$ where the χ^2's have independent χ^2-distributions with degrees of freedom as indicated in the second subscripts and

$$C_1 = (2n)^{-1}\{1 + [1 - 4\,|\,c\,|/(c_{11} + c_{22})^2]^{\frac{1}{2}}\},$$

$$C_2 = (2n)^{-1}\{1 - [1 - 4\,|\,c\,|/(c_{11} + c_{22})^2]^{\frac{1}{2}}\},$$

and $|\,c\,| = c_{11}c_{22} - c_{12}^2$.

The distribution of a quantity similar to U was first obtained by Hsu [9]. An independent derivation of the distribution of the quantity $\tau = \xi(\lambda_1\chi_1^2 + \lambda_2\chi_2^2)^{-\frac{1}{2}}$, where λ_1, λ_2 are certain positive constants and ξ is $N(0, 1)$, will also be found in [10]. Robbins and Pitman [11] have obtained general results for the distribution of the ratio of mixtures of χ^2's, of which the form (4.2.1) given below is a particular case.

We have the following two forms for the frequency function of U:

$$g_0(U) = [B(n, \tfrac{1}{2})U^{\frac{1}{2}}]^{-1}(C_2/C_1)^{n/2}C_2^{\frac{1}{2}}(1 + C_2U)^{-n-\frac{1}{2}} \cdot F\left(n + \tfrac{1}{2}, \tfrac{1}{2}n, n, \frac{1 - C_2/C_1}{1 + C_2U}\right) \tag{4.2.1}$$

for any value of n, and

$$\begin{aligned} g_0(U) = U^{-\frac{1}{2}} \sum_{h=0}^{\frac{1}{2}n-1} \Big(\Gamma(\tfrac{1}{2}n)\Gamma(\tfrac{1}{2})\Gamma(\tfrac{1}{2}n - h)\Gamma(h + 1)\Big)^{-1}\Gamma(\tfrac{1}{2}(n + 1) - h) \\ \cdot\Gamma(\tfrac{1}{2}n + h)(C_1 - C_2)^{-\frac{1}{2}n-h} \\ \cdot[(-1)^hC_1^{\frac{1}{2}(n+1)}C_2^h(1 + C_1U)^{-\frac{1}{2}(n+1)+h} + (-1)^{\frac{1}{2}n}C_2^{\frac{1}{2}(n+1)}C_1^h(1 + C_2U)^{-\frac{1}{2}(n+1)+h}], \end{aligned} \tag{4.2.2}$$

for n even [10].

We notice that since $C_1 + C_2 = 1/n$, the distribution of U essentially depends on C_1 or C_2 and is, therefore, precisely determined by n and the quantity $c/(\mathrm{tr}\ c)^2$ $(=a/(\mathrm{tr}\ a)^2) = w$ (say). If $w > \frac{1}{4}$, C_1 and C_2 are both imaginary. When the matrix c is a 2×2 matrix, the truth of the relation $0 \leq w \leq \frac{1}{4}$ can also be verified independently. The relation (4.2.1) is not defined when $C_2 = 0$, i.e., when $w = 0$; however it is clear from the form of U that it is distributed as Student's t^2 with n degrees of freedom. If $w = \frac{1}{4}$, $C_1 = C_2 = 1/(2n)$, and U has the t^2 distribution with $2n$ degrees of freedom. We shall refer to this again in the

next section where we examine the overall behavior of the probability of Type I error of U with respect to w.

4.3. *Probability of Type I error of U.* To derive $P = P(U \geq U_0)$ corresponding to the form of the frequency function (4.2.1) we put $\zeta = (1 + C_2 U)^{-1}$ and after integration obtain

$$P = (C_2/C_1)^{\frac{1}{2}n} \sum_{h=0}^{\infty} \Gamma(\tfrac{1}{2}n + h)[\Gamma(\tfrac{1}{2}n)\Gamma(h+1)]^{-1}(1 - C_2/C_1)^h \cdot I_{\zeta_0}(n + h, \tfrac{1}{2}), \tag{4.3.1}$$

where $I_{\zeta_0}(p, q)$ is the incomplete beta ratio and $\zeta_0 = (1 + C_2 U_0)^{-1}$. The series (4.3.1) consists of positive terms and is absolutely and uniformly convergent. Corresponding to the form (4.2.2) for even degrees of freedom we similarly obtain

$$P = \sum_{h=0}^{\frac{1}{2}n-1} \Gamma(\tfrac{1}{2}n + h)(\Gamma(\tfrac{1}{2}n)\Gamma(h+1))^{-1}(C_1 - C_2)^{-\frac{1}{2}n-h} \cdot[(-1)^h C_1^{\frac{1}{2}n} C_2^h I_{\zeta_0'}(\tfrac{1}{2}n - h, \tfrac{1}{2}) + (-1)^{\frac{1}{2}n} C_2^{\frac{1}{2}n} C_1^h I_{\zeta_0}(\tfrac{1}{2}n - h, \tfrac{1}{2})], \tag{4.3.2}$$

where $\zeta_0' = (1 + C_1 U_0)^{-1}$.

Consider the series (4.3.1). Following Robbins and Pitman [11] if we set

$$d_h = (C_2/C_1)^{\frac{1}{2}n}\Gamma(\tfrac{1}{2}n + h)(1 - C_2/C_1)^h[\Gamma(\tfrac{1}{2}n)\Gamma(h+1)]^{-1},$$

so that $\Sigma_0^{\infty} d_h = 1$, we have

$$0 \leq P - \sum_{h=0}^{p} d_h I_{\zeta_0}(n + h, \tfrac{1}{2}) \leq \left(1 - \sum_{h=0}^{p} d_h\right) I_{\zeta_0}(n + 2(p+1), \tfrac{1}{2}).$$

For any given U_0 this inequality sets an upper bound to the error committed in P in stopping at the $(p+1)$st term of the series (4.3.1) which has been found to be slowly convergent. Whenever n is even and not large, the finite form (4.3.2) is to be preferred for computational purposes.

We now state the following theorem concerning the dependence of the probability of Type I error of U on the variable parameter w:

THEOREM III. *For any n and fixed U_0, $P(U \geq U_0 \mid H_0)$ is a monotone decreasing function of the variable parameter w.*

PROOF. We shall prove this result by considering the derivative of P with respect to w. From (4.3.1) we obtain

$$\frac{dP}{dw} = (C_2/C_1)^{\frac{1}{2}n-1}(1 - 4w)^{-\frac{1}{2}} \sum_{h=0}^{\infty} \Gamma(\tfrac{1}{2}n + h)[\Gamma(\tfrac{1}{2}n)\Gamma(h + 1]^{-1}$$

$$\cdot\Bigg[4(1 + (1 - 4w)^{\frac{1}{2}})^{-2} I_{\zeta_0}(n + h, \tfrac{1}{2})\{\tfrac{1}{2}n(1 - C_2/C_1)^h - h(C_2/C_1)(1 - C_2/C_1)^{h-1}\}$$

$$- 2\,\frac{C_2/C_1(1 - C_2/C_1)^h \zeta_0^{n+h}(1 - \zeta_0)^{\frac{1}{2}}(1 - (1 - 4w)^{\frac{1}{2}})^{-1}}{\mathrm{B}(n + h, \tfrac{1}{2})}\Bigg],$$

which may actually be shown to represent a derivative. Following Hsu ([9], pp. 14–15) the series

$$\sum\Gamma(\tfrac{1}{2}n + h)(\Gamma(\tfrac{1}{2}n)\Gamma(h + 1))^{-1}[\tfrac{1}{2}n(1 - C_2/C_1)^h - h(C_2/C_1)(1 - C_2/C_1)^{h-1}] \cdot I_{\zeta_0}(n + h, \tfrac{1}{2})$$

can be shown to be equivalent to

$$\sum(\tfrac{1}{2}n + h)(\Gamma(\tfrac{1}{2}n)\Gamma(h + 1))^{-1}\Gamma(\tfrac{1}{2}n + h)(1 - C_2/C_1)^h\eta_h,$$

where

$$\eta_h = (n + h)^{-1}\zeta_0^{n+h}(1 - \zeta_0)^{\frac{1}{2}}/B(n + h, \tfrac{1}{2}) = I_{\zeta_0}(n + h, \tfrac{1}{2}) - I_{\zeta_0}(n + h + 1, \tfrac{1}{2}).$$

After some simplification we obtain

$$\frac{dP}{dw} = (1 - 4w)^{-\frac{1}{2}}(C_2/C_1)^{\frac{1}{2}n-1}(nC_1)^{-2} \sum_{h=0}^{\infty} \Gamma(\tfrac{1}{2}n + h)(\Gamma(\tfrac{1}{2}n)\Gamma(h + 1))^{-1}(1 - C_2/C_1)^h \cdot[\tfrac{1}{2}n(1 - 2nC_1) + h(1 - nC_1)]\eta_h. \tag{4.3.3}$$

The terms of the series (4.3.3) will be negative in the beginning but will finally become positive. Let the $(r + 1)$st term be the first positive term. Since η_h is a monotone decreasing function of h we have

$$\begin{aligned}\frac{dP}{dw} &< \eta_r(1 - 4w)^{-\frac{1}{2}}(C_2/C_1)^{\frac{1}{2}n-1}(nC_1)^{-2} \cdot\sum_{h=0}^{\infty} \Gamma(\tfrac{1}{2}n + h)(\Gamma(\tfrac{1}{2}n)\Gamma(h + 1))^{-1}(1 - C_2/C_1)^h \cdot[\tfrac{1}{2}n(1 - 2nC_1) + h(1 - nC_1)] \\ &= \eta_r(1 - 4w)^{-\frac{1}{2}}(C_2/C_1)^{\frac{1}{2}n-1}(2nC_1^2)^{-1} \cdot[(1 - 2nC_1)(C_2/C_1)^{-\frac{1}{2}n} + (1 - C_2/C_1)(1 - nC_1)(C_2/C_1)^{-\frac{1}{2}n-1}] \\ &= 0.\end{aligned}$$

This proves the theorem except for the end point $w = 0$ of the interval $0 \leq w \leq \frac{1}{4}$, for which the series (4.2.1) and consequently (4.3.1) are not defined. To cover this point we need only to note that the cumulative distribution function (cdf) of the statistic $U = \chi^2_{1,1}((1/n)\chi^2_{2,n} + C_2(\chi^2_{3,n} - \chi^2_{2,n}))^{-1}$ is a continuous function of C_2 and that when $C_2 \to 0$, the cdf of U tends to the cdf of Student's t^2 for n degrees of freedom.

Having established the monotone nature of P with respect to w we are now in a position to assert that U could be regarded as Student's t^2 with degrees of freedom lying between n and $2n$.

4.4 *A p-variate generalization of U.* The reader will at once recognize the following technique for obtaining the generalization of U to the p-variate case to be similar to that of obtaining Hotelling's T from the ordinary Student's t. Consider

$$B_{12} = \alpha_1 b_{12} + \alpha_2 b_{13} + \alpha_3 b_{23} + \cdots + \alpha_{\frac{1}{2}p(p-1)} b_{p-1,p},$$
$$B_{21} = \alpha_1 b_{21} + \alpha_2 b_{31} + \alpha_3 b_{32} + \cdots + \alpha_{\frac{1}{2}p(p-1)} b_{p,p-1}.$$

Let A denote the sample covariance matrix of the $\frac{1}{2}p(p-1)$ symmetric differences. Define row vectors

$$\alpha = (\alpha_1, \alpha_2, \cdots), \qquad b_1 = (b_{12}, b_{13}, \cdots), \qquad b_2 = (b_{21}, b_{31}, \cdots)$$

and let α', b_1', b_2' denote the corresponding column vectors. If we regard α's as constants, then

$$\mathfrak{U} = \frac{(B_{12} - B_{21})^2}{\text{Estimated var}(B_{12} - B_{21})} = \frac{[\alpha(b_1 - b_2)']^2}{\alpha A \alpha'}$$

is also distributed like U. If we determine α's so as to maximize $\mathfrak{U}$ we at once find that $\alpha \propto (b_1 - b_2)A^{-1}$ and we have

$$\mathfrak{U} = (b_1 - b_2)A^{-1}(b_1 - b_2)',$$

which reduces to U when $p = 2$. It can be shown after a very laborious simplification that $\mathfrak{U}$ is also invariant under all contragredient transformations. The distribution of $\mathfrak{U}$ is still under investigation.

4.5. *The power function of U and its unbiased character.* If we let

$$\delta = (\beta_{12} - \beta_{21})(c_{22}\sigma_{11} + c_{11}\sigma_{22} - 2c_{12}\sigma_{12})^{-\frac{1}{2}},$$

we shall presently see that except for the noncentral χ^2 in the numerator the non null distribution of U is similar to its null distribution. We shall first indicate the results that can at most be accomplished in the non null case by the contragredient transformation of y and x. Noting that under this type of transformation β and σ are transformed cogrediently we state:

LEMMA 1. *If $\beta_{12} \neq \beta_{21}$, there does not exist a nonsingular cogredient transformation f which will reduce β to a diagonal matrix and σ to an identity matrix.*

PROOF. Suppose there exists an f such that

$$f\sigma f' = I, \qquad f\beta f' = D,$$

where I is an identity and D a diagonal matrix. Therefore $\beta = f^{-1}Df'^{-1}$; $\beta' = f^{-1}Df'^{-1}$, yielding $\beta = \beta'$, which is contrary to the hypothesis.

LEMMA 2. *If $\beta_{12} \neq \beta_{21}$, there exists a nonsingular transformation f which reduces σ to $\sigma^* = \begin{Vmatrix} 1 & \rho \\ \rho & 1 \end{Vmatrix}$, β to another nonsymmetric matrix β^* and which leaves the standardised "distance" δ between the two alternatives invariant.*

PROOF. Such a transformation is given by $f = \begin{Vmatrix} \sigma_{11}^{-\frac{1}{2}} & 0 \\ 0 & \sigma_{22}^{-\frac{1}{2}} \end{Vmatrix}$. This completes the proof.

We may thus derive the non null distribution of U assuming $\sigma_{11} = \sigma_{22} = 1$. We shall presently see that the power function of U depends only on one nuisance parameter ρ.

To reduce the positive definite form Z in the denominator of U to a linear combination of two independently distributed χ^2's we proceed as follows:

(i) There exists an orthogonal transformation which will simultaneously yield $s_{11} = \Sigma_1^n z_{1\alpha}^2/n$; $s_{22} = \Sigma_1^n z_{2\alpha}^2/n$; $s_{12} = \Sigma_1^n z_{1\alpha} z_{2\alpha}/n$, where $z_{1\alpha}$ and $z_{2\alpha}$ follow a certain bivariate law.

(ii) The transformation

$$z_{1\alpha}^* = (1 - \rho^2)^{-\frac{1}{2}}(z_{1\alpha} - \rho z_{2\alpha})$$

$$z_{2\alpha}^* = z_{2\alpha}$$

further reduces Z to a quadratic form in normally and independently distributed variates.

(iii) A proper choice of θ in the orthogonal transformation

$$z_{1\alpha}' = z_{1\alpha}^* \cos\theta - z_{2\alpha}^* \sin\theta$$

$$z_{2\alpha}' = z_{1\alpha}^* \sin\theta + z_{2\alpha}^* \cos\theta$$

ensures the vanishing of sample covariance of $z_{1\alpha}'$ and $z_{2\alpha}'$ and we obtain $nz = q_1\Sigma_1^n z_{1\alpha}'^2 + q_2\Sigma_1^n z_{2\alpha}'^2$, where q_1 and q_2 depend upon ρ and the elements of the matrix c.

Finally we have $U = \chi_{1,1}'^2(\gamma_1\chi_{2,n}^2 + \gamma_2\chi_{3,n}^2)^{-1}$, where χ'^2 is a noncentral χ^2 and

$$\gamma_1 = (2n)^{-1}[1 + (1 - 4\,|\,c\,|(1 - \rho^2)(c_{11} + c_{22} - 2\rho c_{12})^{-2})^{\frac{1}{2}}],$$

$$\gamma_2 = (2n)^{-1}[1 - (1 - 4\,|\,c\,|(1 - \rho^2)(c_{11} + c_{22} - 2\rho c_{12})^{-2})^{\frac{1}{2}}].$$

We observe that if the covariance matrix is an identity matrix, the values of γ_1 and γ_2 check with the values of C_1 and C_2 (Section 4.3).

Following Hsu [9] we obtain the following forms for the non null frequency function and power function of U:

$$g(U) = e^{-\frac{1}{2}\delta^2}(\gamma_1/\gamma_2)^{\frac{1}{2}n} \sum_{r=0}^{\infty} (\tfrac{1}{2}\delta^2)^r \frac{\gamma_2^{r+\frac{1}{2}} U^{r-\frac{1}{2}}(1 + \gamma_2 U)^{-n-r-\frac{1}{2}}}{\Gamma(r+1)B(n, r+\frac{1}{2})} \cdot F\left(n + r + \tfrac{1}{2}, \tfrac{1}{2}n, n, \frac{1 - \gamma_2/\gamma_1}{1 + \gamma_2 U}\right) \tag{4.5.1}$$

and

$$\beta(\delta, \rho, n) = e^{-\frac{1}{2}\delta^2}(\gamma_2/\gamma_1)^{\frac{1}{2}n} \sum_{h=0}^{\infty}\sum_{r=0}^{\infty} (\tfrac{1}{2}\delta^2)^r\Gamma(\tfrac{1}{2}n + h)(1 - \gamma_2/\gamma_1)^h \cdot [\Gamma(\tfrac{1}{2}n)\Gamma(r+1)\Gamma(h+1)]^{-1} I_{a_0}(n + h, r + \tfrac{1}{2}) \tag{4.5.2}$$

where F denotes the hypergeometric function and $a_0 = (1 + \gamma_2 U_0)^{-1}$. Because of the fixed relation $\gamma_1 + \gamma_2 = 1/n$ either of the above two results could be expressed in terms of γ_1 or γ_2 and consequently ρ is the only nuisance parameter present in (4.5.1) and (4.5.2).

To show that U provides an unbiased test for the hypothesis $\beta_{12} = \beta_{21}$ we state the following theorem:

THEOREM IV. *For any n and fixed ρ the power function $\beta(\delta, \rho, n)$ is a monotone increasing function of the standardised "distance" δ beween the two alternatives.*

PROOF. Consider the double series

$$\sum_{r=0}^{\infty} \sum_{h=0}^{\infty} (\tfrac{1}{2}\delta^2)^r \Gamma(\tfrac{1}{2}n + h)(1 - \gamma_2/\gamma_1)^h [\Gamma(\tfrac{1}{2}n)\Gamma(r + 1)\Gamma(h + 1)]^{-1} I_{a_0}(n + h, r + \tfrac{1}{2})$$

which is dominated by

$$\sum_{r=0}^{\infty} (\gamma_1/\gamma_2)^{\frac{1}{2}n} (\tfrac{1}{2}\delta^2)^r / r!$$

This latter series has infinite radius of convergence and consequently we can differentiate (4.5.2) term by term. Setting $\frac{1}{2}\delta^2 = \Delta^*$ and differentiating we obtain after simplification

$$\frac{\partial\beta(\delta, \rho, n)}{\partial\Delta^*} = (\gamma_2/\gamma_1)^{\frac{1}{2}n} e^{-\Delta^*} \sum_{h=0}^{\infty} \sum_{r=0}^{\infty} \Gamma(\tfrac{1}{2}n + h)(1 - \gamma_2/\gamma_1)^h \Delta^{*r}$$
$$\cdot (\Gamma(\tfrac{1}{2}n)\Gamma(h + 1)\Gamma(r + 1))^{-1} [I_{a_0}(n + h, r + \tfrac{3}{2}) - I_{a_0}(n + h, r + \tfrac{1}{2})].$$

Since $I_{a_0}(n + h, r + \frac{3}{2}) - I_{a_0}(n + h, r + \frac{1}{2}) > 0$, therefore $\partial\beta(\delta, \rho, n)/\partial\Delta^* > 0$. This proves the theorem and establishes the unbiased character of the test based on U.

REFERENCES

[1] H. HOTELLING, "Edgeworth's taxation paradox and the nature of demand and supply functions," *Jour. Polit. Economy*, Vol. 40 (1932), pp. 577–616.

[2] H. SCHULTZ, *The Theory and Measurement of Demand*, University of Chicago Press, 1938, ch. 18.

[3] M. BOCHER, *Introduction to Higher Algebra*, The Macmillan Company, New York, 1907.

[4] H. HOTELLING, "Relations between two sets of variates," *Biometrika*, Vol. 28 (1936), pp. 321–377.

[5] H. HOTELLING, "Spaces of statistics and their metrization," *Science*, Vol. 67 (1928), pp. 149–150.

[6] H. HOTELLING, "Commodity transformations and matrices," *Annals of Math. Stat.*, Vol. 10 (1939), p. 88.

[7] J. WISHART, "The generalized product-moment distribution in samples from a normal multivariate population," *Biometrika*, Vol. 20 (1928), pp. 32–52.

[8] S. S. WILKS, *Mathematical Statistics*, Princeton University Press, 1946, pp. 245–247.

[9] P. L. HSU, "Contribution to the theory of Student's t-test as applied to the problem of two samples," *Stat. Res. Memoirs*, Vol. 2 (1938), pp. 1–24.

[10] UTTAM CHAND, "Distributions related to the comparison of two means and two regression coefficients," *Annals of Math. Stat.*, Vol. 21 (1950), pp. 507–522.

[11] H. ROBBINS AND E. J. G. PITMAN, "Application of the method of mixtures to quadratic forms in normal variates," *Annals of Math. Stat.*, Vol. 20 (1949), pp. 552–560.

[12] TJALLING C. KOOPMANS, ed., *Statistical Inference in Dynamic Economic Models*, Cowles Commission Monograph 10, John Wiley and Sons, New York, 1950

EXTREMAL PROPERTIES OF EXTREME VALUE DISTRIBUTIONS

By Sigeiti Moriguti

University of North Carolina

Summary. The upper and lower bounds for the expectation, the coefficient of variation, and the variance of the largest member of a sample from a symmetric population are discussed. The upper bound for the expectation (Table 1, Fig. 1), the lower bound for the C.V. (Table 2, Fig. 4) and the lower bound for the variance (Fig. 7) are actually achieved for the corresponding particular population distributions (Figs. 2, 3, 5, 6, equation (5.1)). The rest of the bounds are not actually achieved but approached as the limits, for example, for the three-point distribution (Section 3) by letting p tend to zero.

1. Introduction. The sampling distribution of the largest or the smallest member of a sample has been studied by several authors; Tippett [1] and de Finetti [2] considered a sample from a normal population, Olds [3] from a rectangular population. The case of a very large sample was treated by Dodd [4], Fisher and Tippett [5], and Gumbel [6], each for a certain class of population distributions.

Here we consider the upper and lower bounds for the expectation, the coefficient of variation, and the variance of the extreme member of a sample from a symmetrically distributed population with a finite variance. To be specific, we will discuss only the largest member and take the mean of the population equal to zero. These conventions do not imply any essential restriction.

2. Notations and formulas. Let the cumulative distribution function (cdf) of the population be denoted by $F(x)$; then the cdf of the largest member x_n of a sample of size n is given by $\{F(x)\}^n$. Hence the expectation of the largest member can be expressed by

$$E(x_n) = \int_{-\infty}^{\infty} xn\{F(x)\}^{n-1}\, dF(x). \tag{2.1}$$

Now we consider the inverse function $x(F)$ of $F(x)$, with an obvious additional definition at points of discontinuity, if any, of $F(x)$. Thus (2.1) can also be written as

$$E(x_n) = \int_0^1 x(F)nF^{n-1}\, dF. \tag{2.2}$$

Because of symmetry, $x(F) = -x(1 - F)$ holds almost everywhere, whence

$$E(x_n) = \int_{\frac{1}{2}}^1 x(F)n\{F^{n-1} - (1 - F)^{n-1}\}\, dF. \tag{2.3}$$

Similarly, we get as the variance

$$V(x_n) = \int_{\frac{1}{2}}^{1} \{x(F)\}^2 n\{F^{n-1} + (1 - F)^{n-1}\}\, dF - \{E(x_n)\}^2. \tag{2.4}$$

The population variance is of course given by

$$\sigma^2 = 2 \int_{\frac{1}{2}}^{1} \{x(F)\}^2\, dF. \tag{2.5}$$

3. Bounds for the expectation of the largest member. In Schwarz's inequality

$$\left(\int_a^b f(F)g(F)\, dF\right)^2 \leqq \int_a^b \{f(F)\}^2\, dF \int_a^b \{g(F)\}^2\, dF, \tag{3.1}$$

putting $a = \frac{1}{2}$, $b = 1$, $f(F) \equiv x(F)$, $g(F) \equiv n\{F^{n-1} - (1 - F)^{n-1}\}$, we get a formula which means, in view of (2.3) and (2.5), that

$$E(x_n) \leqq \frac{\sigma}{\sqrt{2}}\, n \left(\int_{\frac{1}{2}}^{1} \{F^{n-1} - (1 - F)^{n-1}\}^2\, dF\right)^{\frac{1}{2}}, \tag{3.2}$$

equality being satisfied if and only if $f(F) = \text{const.}\cdot g(F)$, that is,

$$x(F) = \text{const.}\cdot \{F^{n-1} - (1 - F)^{n-1}\}. \tag{3.3}$$

Therefore the expectation of the largest member has the right-hand side of (3.2) as an upper bound, which is actually achieved for a particular type of population distribution given by (3.3).

The integral in (3.2) can easily be evaluated as follows:

$$\begin{aligned} &\int_{\frac{1}{2}}^{1} \{F^{n-1} - (1 - F)^{n-1}\}^2\, dF \\ &= \frac{1}{2}\int_0^1 [F^{2n-2} + (1 - F)^{2n-2} - 2F^{n-1}(1 - F)^{n-1}]\, dF \\ &= \frac{1}{2}\left[\frac{1}{2n - 1} + \frac{1}{2n - 1} - 2\mathrm{B}(n, n)\right] = \frac{1}{2n - 1} - \mathrm{B}(n, n), \end{aligned} \tag{3.4}$$

where the Beta function of equal integral arguments can also be expressed as

$$\mathrm{B}(n, n) = \frac{1}{(2n - 1)C_{n-1}^{2n-2}}. \tag{3.5}$$

Thus the upper bound for $E(x_n)$ is given by

$$E(x_n) \leqq \frac{n}{\sqrt{2(2n - 1)}} \left(1 - \frac{1}{C_{n-1}^{2n-2}}\right)^{\frac{1}{2}} \sigma. \tag{3.6}$$

The numerical value of the coefficient is calculated for various sample sizes and compared with the values of $E(x_n)/\sigma$ for normal and rectangular populations in Table 1 and Fig. 1. It is to be noted that the value for a normal population is remarkably close to the upper bound if $n \leqq 7$. The cumulative distribution

curve and frequency curve of the extremal distribution (3.3) is illustrated in Figs. 2 and 3 for several values of sample size n.

It is obvious that the expectation of the largest member has the lower bound zero. However it may be of some interest to see that this lower bound can be approached as closely as one desires. One of the simplest ways is to consider the three-point distribution, such as the values a, 0 and $-a$ occurring with probabilities p, $1 - 2p$, and p, respectively. If we make p approach zero for a fixed sample size n, the ratio $E(x_n)/\sigma$ also approaches zero, because in this case

TABLE 1

Expectation of the largest member in the unit of σ, $E(x_n)/\sigma$

Sample size n	Upper bound	For normal distribution*	For rectangular distribution
2	.5774	.5642	.5774
3	.8660	.8463	.8660
4	1.0420	1.0294	1.0392
5	1.1701	1.1630	1.1547
6	1.2767	1.2672	1.2372
7	1.3721	1.3522	1.2990
8	1.4604	1.4236	1.3472
9	1.5434	1.4850	1.3856
10	1.6222	1.5388	1.4171
11	1.6974	1.5864	1.4434
12	1.7693	1.6292	1.4656
13	1.8385	1.6680	1.4846
14	1.9052	1.7034	1.5011
15	1.9696	1.7359	1.5155
16	2.0320	1.7660	1.5283
17	2.0926	1.7939	1.5396
18	2.1514	1.8200	1.5497
19	2.2087	1.8450	1.5588
20	2.2645	1.8673	1.5671

* From [9], p. 165.

$$E(x_n) = nap + O(p^2), \tag{3.7}$$

$$\sigma^2 = 2a^2p. \tag{3.8}$$

4. Bounds for the coefficient of variation of the largest member of a sample. Putting in (3.1) $a = \frac{1}{2}$, $b = 1$, and

$$
\begin{aligned}
(4.1) \qquad f(F) &\equiv x(F)\sqrt{n}\{F^{n-1} + (1 - F)^{n-1}\}^{\frac{1}{2}}, \\
g(F) &\equiv \frac{\sqrt{n}\{F^{n-1} - (1 - F)^{n-1}\}}{\{F^{n-1} + (1 - F)^{n-1}\}^{\frac{1}{2}}},
\end{aligned}
$$

we get a formula which means, in view of (2.3) and (2.4), that

$$(4.2) \qquad \frac{V(x_n)}{E(x_n)^2} \geqq \frac{1}{M_n} - 1,$$

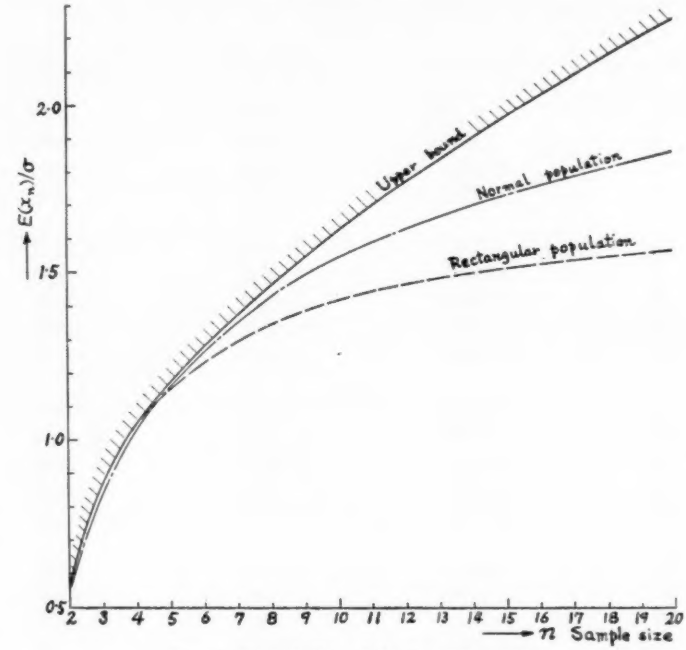

FIG. 1. Expectation of the largest member

where

$$(4.3) \qquad M_n \equiv \int_{\frac{1}{2}}^{1} \frac{n\{F^{n-1} - (1 - F)^{n-1}\}^2}{F^{n-1} + (1 - F)^{n-1}}\, dF.$$

The equality in (4.2) is satisfied if and only if $f = \text{const.}\cdot g$, i.e.

$$(4.4) \qquad x(F) = \text{const.} \cdot \frac{F^{n-1} - (1 - F)^{n-1}}{F^{n-1} + (1 - F)^{n-1}}.$$

Therefore the coefficient of variation of the largest member has $\sqrt{(1/M_n) - 1}$ as a lower bound which is actually achieved for a particular type of population distribution given by (4.4).

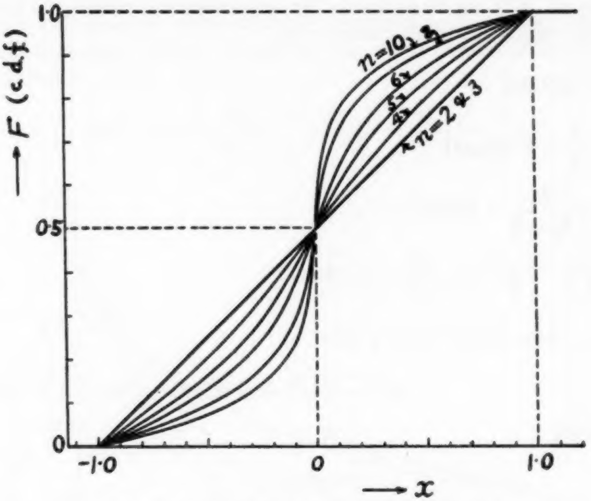

FIG. 2. Extremal distribution (cdf) for $E(x_n)$

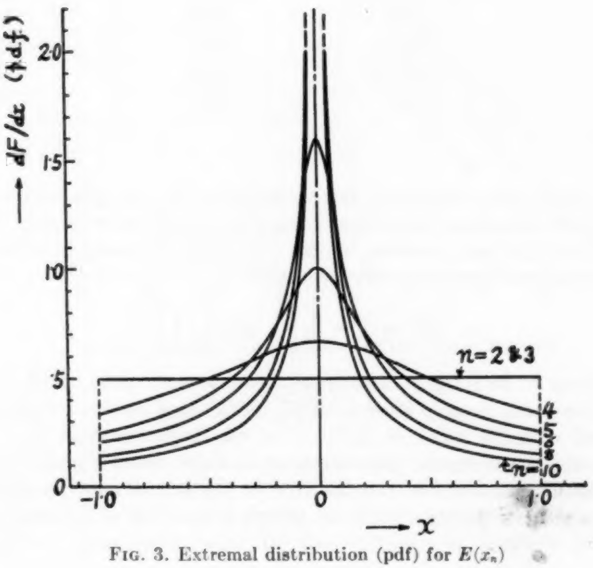

FIG. 3. Extremal distribution (pdf) for $E(x_n)$

The integral (4.3) can be evaluated by an elementary method of quadrature. To show the results for small values of n,

$$M_2 = \frac{1}{3} = 0.33333,$$

$$M_3 = 3 - \frac{3}{4}\pi = 0.64381,$$

$$M_4 = \frac{23}{15} - \frac{32\pi}{81\sqrt{3}} = 0.81677,$$

$$M_5 = -\frac{55}{3} + \frac{35\sqrt{2}}{4}\pi - \frac{25}{4}\pi = 0.90695,$$

$$M_6 = -\frac{6}{7} + 0.6\pi\{(0.704 + 0.8\sqrt{0.8}) \cdot\sqrt{5 - 2\sqrt{5}} + 2(0.704 - 0.8\sqrt{0.8})\sqrt{5 + 2\sqrt{5}}\}$$

$$= 0.95300.$$

TABLE 2

Coefficient of variation of the largest member

Sample size n	Lower bound	For normal population*	For rectangular population
2	1.4142	1.4634	1.4142
3	.7438	.8838	.7746
4	.4737	.6812	.5443
5	.3203	.5752	.4226
6	.2221	.5089	.3464

* Cf. [7].

As the sample size n increases, the evaluation of M_n by quadrature becomes more and more laborious. Numerical integration would be preferable for larger values of n. In this case, however, we can derive (see Appendix 1 for the derivation) an asymptotic formula of M_n for large n

$$M_n = 1 - \frac{\pi}{2^n}\left[1 + O\left(\frac{1}{n}\right)\right] \tag{4.5}$$

which happens to be a fairly close approximation even for as small a value of n as six, where this formula gives 0.95091. Using these results, we compare the lower bound with the value of the C.V. of the largest member for a normal population and a rectangular population, as in Table 2 and Fig. 4.

It is interesting to observe that the C.V. of the largest member of a sample from a two-point population, such as values 1 and -1 each occurring with

probability 1/2, behaves asymptotically similarly to the lower bound except for a numerical factor $\sqrt{\pi}/2$. In fact, we can easily derive in this case the following formulas

$$E(x_n) = 1 - \frac{1}{2^{n-1}}, \qquad V(x_n) = \frac{1}{2^{n-2}} - \frac{1}{2^{2n-2}}, \tag{4.6}$$

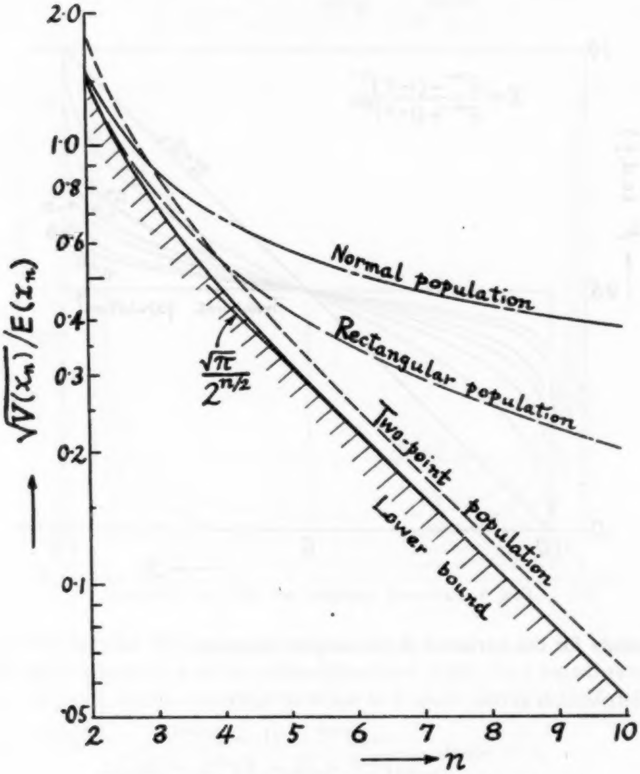

FIG. 4. Coefficient of variation of the largest member

$$\frac{\sqrt{V(x_n)}}{E(x_n)} = \frac{\sqrt{(2^n - 1)}}{2^{n-1} - 1} \approx \frac{1}{2^{\frac{1}{2}n-1}}. \tag{4.7}$$

This similarity in the asymptotic behavior may be taken to be the reflection of the similarity in the population distribution, which is seen in comparing Figs. 5 and 6 with the corresponding graphs for the two-point distribution.

There is no finite upper bound for the coefficient of variation of the largest

member. It can be proved, for instance, by observing the behavior in the case of the three-point distribution mentioned in the previous section when p approaches zero for fixed n. In fact, in this case, it is easy to show that

$$V(x_n) = na^2p + O(p^2),$$

$$\frac{\sqrt{V(x_n)}}{E(x_n)} = \frac{1}{\sqrt{np}} + O(\sqrt{p}) \to \infty. \tag{4.8}$$

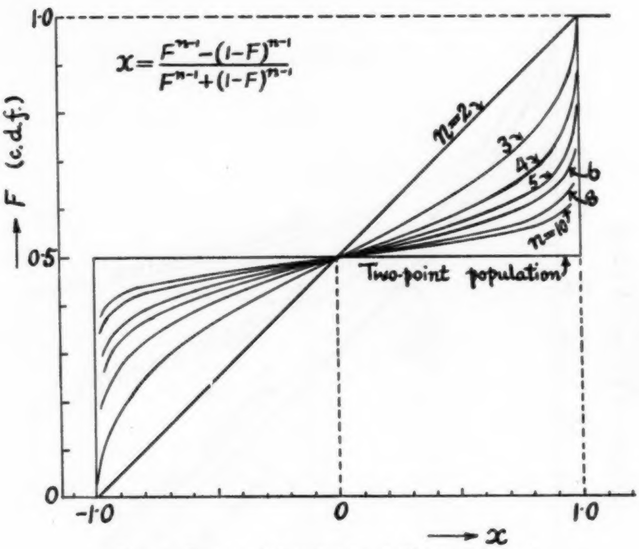

FIG. 5. Extremal distribution (cdf) for C.V.(x_n)

5. Bounds for the variance of the largest member.[1] As we shall prove, $V(x_n)$ has a lower bound $\lambda_n\sigma^2$, which is actually achieved for a particular type of population distribution given, when F is not 0 or 1, by

$$x = \text{const.} \cdot \frac{n\{F^{n-1} - (1-F)^{n-1}\}}{n\{F^{n-1} + (1-F)^{n-1}\} - 2\lambda_n}, \tag{5.1}$$

where λ_n is the only root of the equation[2]

$$M_n(\lambda) \equiv \int_{\frac{1}{2}}^{1} \frac{n^2\{F^{n-2} - (1-F)^{n-1}\}^2}{n\{F^{n-1} + (1-F)^{n-1}\} - 2\lambda}\, dF = 1 \tag{5.2}$$

in the interval $0 \leqq \lambda \leqq n/2^{n-1}$.

---

[1] A heuristic derivation of the formulas (5.1) and (5.2) is given in Appendix 2.

[2] The notation is such that $M_n(0)$ equals M_n as previously defined.

First, in order to prove that there exists one and only one root of (5.2) in the stated interval, it is sufficient to show that

$$M_n(0) < 1, \qquad M_n(n/2^{n-1}) > 1, \tag{5.3}$$

and to note that $M_n(\lambda)$ is a monotone increasing continuous function of λ in the interval. Since

$$\int_{\frac{1}{2}}^{1} n\{F^{n-1} + (1 - F)^{n-1}\}\, dF = 1, \tag{5.4}$$

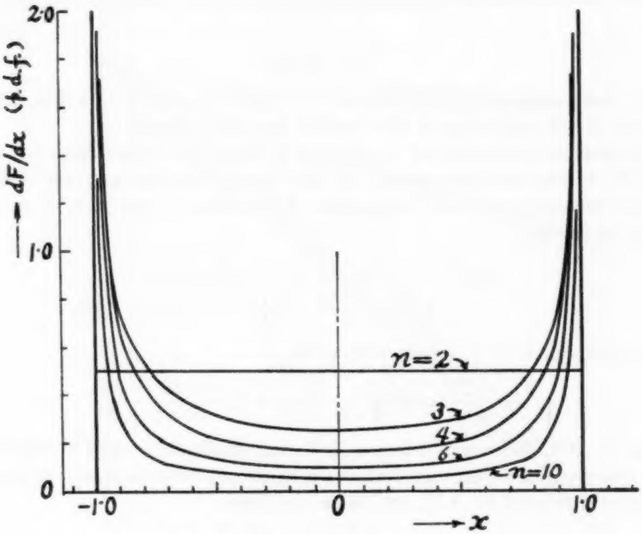

FIG. 6. Extremal distribution (pdf) for C.V.(x_n)

we have, for any λ in the interval,

$$\begin{aligned} 1 - M_n(\lambda) &= \int_{\frac{1}{2}}^{1} \frac{\begin{matrix} n^2\{F^{n-1} + (1 - F)^{n-1}\}^2 - 2\lambda n\{F^{n-1} \\ \qquad + (1 - F)^{n-1}\} - n^2\{F^{n-1} - (1 - F)^{n-1}\}^2 \end{matrix}}{n\{F^{n-1} + (1 - F)^{n-1}\} - 2\lambda}\, dF \\ &= \int_{\frac{1}{2}}^{1} \frac{4n^2F^{n-1}(1 - F)^{n-1} - 2\lambda n\{F^{n-1} + (1 - F)^{n-1}\}}{n\{F^{n-1} + (1 - F)^{n-1}\} - 2\lambda}\, dF. \end{aligned} \tag{5.5}$$

On the other hand, it is obvious that $F^{n-1} + (1 - F)^{n-1}$ assumes a minimum value $1/2^{n-2}$, and $F^{n-1}(1 - F)^{n-1}$ a maximum value $1/2^{2n-2}$ at $F = 1/2$. There-

fore, in the interval[3] $1/2 < F < 1$, the denominator of the integrand is always positive, the numerator being always positive for $\lambda = 0$ and always negative for $\lambda = n/2^{n-2}$. Hence we get (5.3). The above mentioned nature of $M_n(\lambda)$ is also obvious (cf. the definition (5.2) and the above statement about the denominator).

Next, again in the Schwarz's inequality, let us put $a = 1/2$, $b = 1$ and

$$f(F) \equiv x(F)[n\{F^{n-1} + (1 - F)^{n-1}\} - 2\lambda_n]^{\frac{1}{2}}, \tag{5.6}$$

$$g(F) \equiv \frac{n\{F^{n-1} - (1 - F)^{n-1}\}}{[n\{F^{n-1} + (1 - F)^{n-1}\} - 2\lambda_n]^{\frac{1}{2}}}. \tag{5.7}$$

Then we obtain a formula which means, in view of (2.3), (2.4), (2.5) and $M_n(\lambda_n) = 1$, that

$$V(x_n) \geqq \lambda_n \sigma^2, \tag{5.8}$$

equality being satisfied if and only if $f = \text{const.}\cdot g$, i.e. (5.1) holds. Thus the statement at the beginning of this section has been proved.

The numerical evaluation of λ_n requires a little more effort than the evaluation of M_n in the previous section, as the former requires solution of a transcendental equation after an integration. For instance,[4] for $n = 3$, λ_3 can be obtained by solving

$$\tan^{-1} \frac{1}{\sqrt{1 - \frac{4}{3}\lambda}} = \frac{2}{3\sqrt{1 - \frac{4}{3}\lambda}},$$

as $\lambda_3 = .394$. For $n = 4$, we have to solve

$$\sqrt{\frac{1 - 2\lambda}{3}} \tan^{-1} \sqrt{\frac{3}{1 - 2\lambda}} = 1 - \frac{88 - 5\lambda}{10(4 + \lambda)^2}$$

to get $\lambda_4 = .209$. Moreover, when $n \geqq 7$, the quadrature itself is tedious. For large n, however, an asymptotic formula is again available as shown in Appendix 3. It is closely related to (4.5), and takes the form

$$\lambda_n = \frac{\pi}{2^n}\left[1 + O\left(\frac{1}{n}\right)\right]. \tag{5.9}$$

Again it is fairly close even if n is small.

The general picture is seen in Fig. 7, in which the lower bound of $\sqrt{V(x_n)}/\sigma$ is shown together with the value for normal [7], rectangular, and two-point distributions.

As for the upper bound, it is easy to see, from (2.4) and (2.5), that

$$V(x_n) < n \int_{\frac{1}{2}}^{1} x(F)^2\, dF = \tfrac{1}{2} n\sigma^2, \tag{5.10}$$

[3] The suspicion about the singularity which might occur in the case of $\lambda = n/2^{n-2}$ at $F = \frac{1}{2}$ is dissolved if we note that the numerator also has a zero of the second order at $F = \frac{1}{2}$.

[4] For $n = 2$, (5.1) reduces to a rectangular distribution, for which no more calculation is necessary.

for $F^{n-1} + (1 - F)^{n-1}$ is a monotone increasing function taking the value unity at the end $F = 1$ of the interval. The value $n/2$ of the ratio $V(x_n)/\sigma^2$ can be

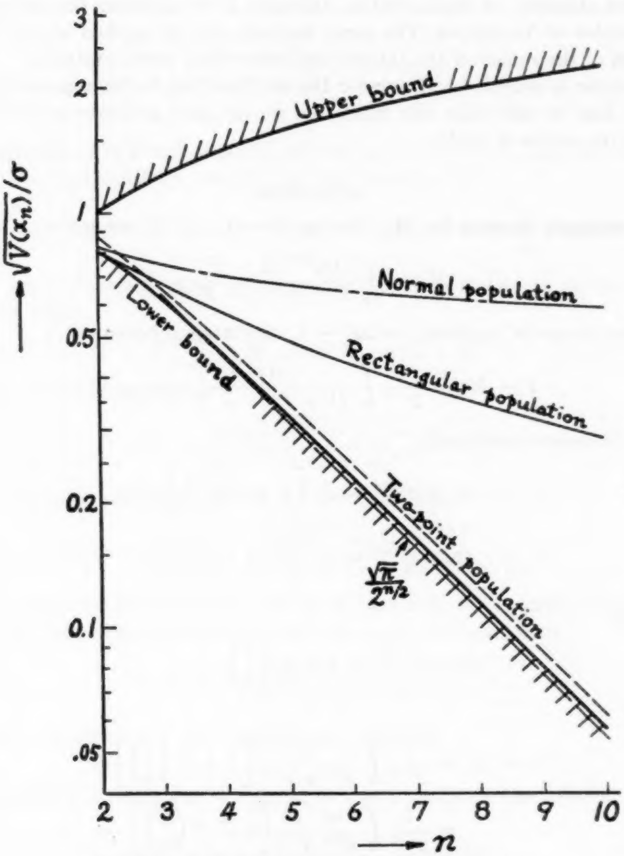

FIG. 7. Standard deviation of the largest member

approached as closely as desired, for example, for the three-point distribution (Section 3) by letting p be sufficiently small. (See (3.8) and (4.8).)

6. Final remarks and acknowledgement. We considered the upper and lower bounds for the expectation, the coefficient of variation, and the variance of the largest member of a sample from a symmetric population. The upper bound for

the expectation and the lower bound for the C.V. or the variance are actually achieved for particular distributions, which we may call "extremal distributions". These distributions as well as the values of the corresponding bounds were first obtained, as illustrated in Appendix 2, by applying the techniques of the Calculus of Variations. The same methods can be applied also to the distribution of the range[5] of the sample and some other useful statistics.

The writer is indebted to Professor Harold Hotelling for his suggestions which induced him to undertake this study and for his kind guidance and encouragement in the course of study.

APPENDICES

1. Asymptotic formula for M_n. Putting $\lambda = 0$ in (5.5), we get

$$1 - M_n = \int_{\frac{1}{2}}^{1} \frac{4nF^{n-1}(1-F)^{n-1}}{F^{n-1} + (1-F)^{n-1}}\, dF.$$

With the change of variable $t = 2F - 1$, this integral becomes

$$1 - M_n = \frac{1}{2^{n-1}} \int_0^1 \frac{n(1-t^2)^{n-1}}{(1+t)^{n-1} + (1-t)^{n-1}}\, dt.$$

When n increases indefinitely,

$$(1+t)^{n-1} = e^{nt}\left[1 + O\left(\frac{1}{n}\right)\right],$$

$$(1-t)^{n-1} = e^{-nt}\left[1 + O\left(\frac{1}{n}\right)\right];$$

therefore,

$$(1-t^2)^{n-1} = 1 + O\left(\frac{1}{n}\right).$$

Thus,

$$\begin{aligned} 1 - M_n &= \frac{1}{2^{n-2}} \int_0^1 \frac{n}{e^{nt} + e^{-nt}} \left[1 + O\left(\frac{1}{n}\right)\right] dt \\ &= \frac{1}{2^{n-2}} \int_0^1 \frac{ne^{nt}\, dt}{e^{2nt} + 1} \left[1 + O\left(\frac{1}{n}\right)\right] \\ &= \frac{1}{2^{n-2}} (\tan^{-1} e^{nt}) \Big|_0^1 \left[1 + O\left(\frac{1}{n}\right)\right] \\ &= \frac{1}{2^{n-2}} \left(\tan^{-1} e^{n} - \frac{\pi}{4}\right)\left[1 + O\left(\frac{1}{n}\right)\right]. \end{aligned}$$

---

[5] Thanks are due to Professor Olds at Carnegie Institute of Technology for calling the author's attention to R. L. Plackett's paper [8] which derived essentially the same result as given in Section 3 of the present paper by a somewhat different approach.

But

$$\tan^{-1} e^n = \frac{\pi}{2} + o\left(\frac{1}{n}\right).$$

Therefore,

$$1 - M_n = \frac{\pi}{2^n}\left[1 + O\left(\frac{1}{n}\right)\right].$$

This is (4.5).

2. Derivation of (5.1) and (5.2). In order to minimize (2.4) under the condition that (2.5) is kept constant, we put the first variation of

$$\int_{\frac{1}{2}}^{1} x(F)^2 n\{F^{n-1} + (1-F)^{n-1}\}\, dF - \{E(x_n)\}^2 - 2\lambda \int_{\frac{1}{2}}^{1} x(F)^2\, dF$$

equal to zero, of course taking account of (2.3). Thus we obtain as the characteristic equation

$$x(F)n\{F^{n-1} + (1-F)^{n-1}\} - E(x_n)n\{F^{n-1} - (1-F)^{n-1}\} - 2\lambda x(F) = 0,$$

which can easily be solved as

$$x(F) = \frac{E(x_n)n\{F^{n-1} - (1-F)^{n-1}\}}{n\{F^{n-1} + (1-F)^{n-1}\} - 2\lambda}.$$

But this solution is eligible only if it satisfies (2.3), that is only if

$$E(x_n) = E(x_n) \int_{\frac{1}{2}}^{1} \frac{n^2\{F^{n-1} - (1-F)^{n-1}\}^2}{n\{F^{n-1} + (1-F)^{n-1}\} - 2\lambda}\, dF.$$

As $E(x_n)$ cannot be zero except in the trivial case $x(F) \equiv 0$, λ must be a solution of (5.2). If there exists a solution λ_n as is actually the case, then

$$x = \text{const.} \cdot \frac{n\{F^{n-1} - (1-F)^{n-1}\}}{n\{F^{n-1} + (1-F)^{n-1}\} - 2\lambda_n}$$

is eligible as a solution of the characteristic equation.

3. Asymptotic formula for λ_n. $M_n(\lambda)$ can be transformed as follows.

$$M_n(\lambda) = \int_{\frac{1}{2}}^{1} \frac{n^2[\{F^{n-1} + (1-F)^{n-1}\} - 2(1-F)^{n-1}]^2}{n\{F^{n-1} + (1-F)^{n-1}\} - 2\lambda}\, dF$$

$$= \int_{\frac{1}{2}}^{1} \Bigg[n\{F^{n-1} + (1-F)^{n-1}\} + 2\lambda - 4n(1-F)^{n-1}$$

$$+ \frac{\{2\lambda - 2n(1-F)^{n-1}\}^2}{n\{F^{n-1} + (1-F)^{n-1}\} - 2\lambda}\Bigg] dF$$

$$= 1 + \lambda - \frac{4}{2^n} + \int_{\frac{1}{2}}^{1} \frac{\{2\lambda - 2n(1-F)^{n-1}\}^2}{n\{F^{n-1} + (1-F)^{n-1}\} - 2\lambda}\, dF.$$

Therefore, λ_n must satisfy the equation

$$\lambda_n = \frac{4}{2^n} - \int_{\frac{1}{2}}^{1} \frac{\{2\lambda_n - 2n(1-F)^{n-1}\}^2}{n\{F^{n-1} + (1-F)^{n-1}\} - 2\lambda_n}\, dF.$$

As the integral is positive, we get $\lambda_n < 4/2^n$. This inequality certifies that the last term of the denominator in the last integral, or in (5.5), can be neglected as of order $1/n$ times that of the first term. Therefore

$$\lambda_n = \int_{\frac{1}{2}}^{1} \frac{4nF^{n-1}(1-F)^{n-1}}{F^{n-1} + (1-F)^{n-1}}\, dF \left[1 + O\left(\frac{1}{n}\right)\right]$$

$$= (1 - M_n)\left[1 + O\left(\frac{1}{n}\right)\right]$$

$$= \frac{\pi}{2^n}\left[1 + O\left(\frac{1}{n}\right)\right].$$

REFERENCES

[1] L. H. C. Tippett, "On the extreme individuals and the range of samples taken from a normal population," *Biometrika*, Vol. 17 (1925), pp. 364–387.

[2] B. de Finetti, "Sulla legge di probabilità degli estremi," *Metron*, Vol. 9 (1932), pp. 127–138.

[3] E. G. Olds, "Distribution of greatest variates, least variates and intervals of variation in samples from a rectangular universe," *Bull. Am. Math. Soc.*, Vol. 41 (1935), pp. 297–304.

[4] E. L. Dodd, "The greatest and the least variate under general laws of error," *Trans. Am. Math. Soc.*, Vol. 25 (1923), pp. 525–539.

[5] R. A. Fisher and L. H. C. Tippett, "Limiting forms of the frequency distribution of the largest or smallest member of a sample," *Proc. Cambridge Philos. Soc.*, Vol. 24 (1928), pp. 180–190.

[6] E. J. Gumbel, "Les valeurs extrêmes des distributions statistiques," *Annales Institut Henri Poincaré*, Vol. 4 (1935), pp. 115–158.

[7] H. J. Godwin, "Some low moments of order statistics," *Annals of Math. Stat.*, Vol. 20 (1949), pp. 279–285.

[8] R. L. Plackett, "Limits of the ratio of mean range to standard deviation," *Biometrika*, Vol. 34 (1947), pp. 120–122.

[9] K. Pearson, *Tables for Statisticians and Biometricians*, Part II, 1st ed., Cambridge University Press, 1931.

THE FITTING OF POLYNOMIALS BY THE METHOD OF WEIGHTED GROUPING

By P. G. Guest

University of Sydney, Australia

Summary. A method of fitting polynomials to equally spaced data is developed which is more rapid than the method of least squares. The orthogonal polynomial $T_j(x)$ of the least squares method is replaced by a step function $w_j(x)$, and this greatly reduces the number of multiplications. An efficiency of about 90 per cent is obtained for the estimates of the coefficients and fitted values.

1. Introduction. An appreciable shortening in the time required to fit a curve to a series of n equally spaced observations $y(x)$ is effected by the use of tables of the orthogonal polynomials $T_j(x)$ or $\xi_j'(x)$ [1], [2], [3]. However, the process is still tedious if the number of observations is at all large. A considerable time is spent in the calculation of the orthogonal moments $\Sigma T_j(x)y(x)$, and a mistake in these calculations can easily be made.

In the present paper a method of curve fitting is developed which considerably reduces the time required for the calculation of the moments. The continuous function $T_j(x)$ is replaced by a step function $w_j(x)$. The observations $y(x)$ are summed over each interval of constancy of $w_j(x)$. The groups so formed are multiplied by the weighting factor $w_j(x)$ and added to give $\Sigma w_j(x)y(x)$.

The number of groups required is found to be $\frac{1}{2}(j + 1)$ or $\frac{1}{2}(j + 2)$ according as j is odd or even. Thus for the coefficients of the fourth and fifth degrees the number of multiplications is reduced to three; the number of weights which have to be tabulated is also reduced to three.

The estimates of the polynomial coefficients and fitted values obtained by the method of grouping are all unbiased, and have an efficiency of about 90 per cent. This means firstly that the standard error of the value obtained by the method of grouping is about 5 per cent greater than the standard error obtained by the method of least squares, and secondly that the probability that the difference between the values obtained by these two methods exceeds their standard error is very small. In practically all cases this efficiency will be quite adequate [4].

A pleasing feature of the method of grouping is that the calculation of the coefficients can be carried out easily without a calculating machine, at least for polynomials of lower degree than the fourth. The calculations can of course be done much more rapidly if a machine is used.

2. Estimation of the power series coefficients. To form an unbiased estimate of the coefficient b_{pj} in the polynomial

$$u_p(x) = \sum_{j=0}^{p} b_{pj} x^j \tag{1}$$

which is to fit the observed values $y(x)$, we must (directly or indirectly) multiply each observation by a weight $w_{pj}(x)$ so chosen that

$$\sum_x w_{pj}(x)x^k = 0, \qquad k \leq p, k \neq j, \tag{2}$$

the sum being taken over the n observations. Then the estimate of b_{pj} is

$$b_{pj} = \sum_x w_{pj}(x)y(x)/\sum_x w_{pj}(x)x^j. \tag{3}$$

The fact that $w_{pj}(x)$ depends not only on j but also on p is a disadvantage. A useful system of weights is obtained by selecting weights $w_{pj}(x)$ which are linear functions of weights $w_j(x)$ independent of p, $w_j(x)$ being chosen so that

$$\sum_x w_j(x)x^k = 0, \qquad k < j. \tag{4}$$

We can then write $w_{pj}(x)$ in the form

$$\frac{w_{pj}(x)}{\Sigma w_j x^j} = \frac{w_j(x)}{\Sigma w_j x^j} + \beta_{j+1,j} \frac{w_{j+1}(x)}{\Sigma w_{j+1} x^{j+1}} + \cdots + \beta_{pj} \frac{w_p(x)}{\Sigma w_p x^p}, \tag{5}$$

where w_j represents $w_j(x)$ and where the coefficients β_{kj} are determined from the condition

$$\Sigma w_{pj}(x)x^k = 0 \qquad j < k \leq p, \tag{6}$$

that is,

$$\frac{\Sigma w_j x^k}{\Sigma w_j x^j} + \beta_{j+1,j} \frac{\Sigma w_{j+1} x^k}{\Sigma w_{j+1} x^{j+1}} + \cdots + \beta_{kj} \frac{\Sigma w_k x^k}{\Sigma w_k x^k} = 0. \tag{7}$$

The advantage of using such a system of weights is that we can introduce statistics

$$a_j = \sum_x w_j(x)y(x)/\sum_x w_j(x)x^j \tag{8}$$

which are independent of p and express b_{pj} as a linear function of these statistics. In fact it follows from equations (3) and (5) that

$$b_{pj} = a_j + \beta_{j+1,j}a_{j+1} + \cdots + \beta_{pj}a_p. \tag{9}$$

The method of least squares is a particular case of this method of weighting [5]. In the method of least squares $w_j(x) = T_j(x)$, the orthogonal polynomial of degree j.

The calculation of the coefficients β_{kj} can be done most conveniently by evaluating the quantities

$$\alpha_{kj} = -\sum_x w_j(x)x^k / \sum_x w_j(x)x^j. \tag{10}$$

Then equation (7) becomes

$$\beta_{kj} = \alpha_{kj} + \alpha_{k,j+1}\beta_{j+1,j} + \cdots + \alpha_{k,k-1}\beta_{k-1,j}, \tag{11}$$

and the coefficients β can be built up in turn from the coefficients α.

3. The method of weighted grouping. In the method of weighted grouping we replace the continuous function $T_j(x)$ occurring in the least squares solution by a step function $w_j(x)$. In effect we assign the same weight $w_j(x)$ to all observations in a region where $T_j(x)$ is fairly constant.

The criterion used in the choice of groups is that of maximum efficiency for the coefficient a_j defined by (8), with $w_j(x)$ satisfying (4). The minimum number of groups required is thus $j + 1$. It would be possible of course to choose a larger number of groups, but this complicates the method without producing any great increase in efficiency. Adopting the value $j + 1$ for the number of groups, the values of the weights for each method of grouping are uniquely determined (except for an arbitrary multiplying factor) by equation (4).

When the observations are equally spaced, it is most convenient to change to a variable ϵ whose origin is at the centre of the points of observation x, and whose scale is such that the interval between successive observations is unity. An obvious simplification of the method of grouping for equally spaced observations is to make the groups symmetrical about the origin; that is, to take $w_j(-\epsilon) = (-)^j w_j(\epsilon)$. This reduces the number of different weights to $\frac{1}{2}(j + 1)$ or $\frac{1}{2}(j + 2)$ according as j is odd or even. Also $\beta_{jk} = 0$ when $j + k$ is odd. The observations are to be grouped by adding corresponding observations $y(\epsilon)$ of equal $|\epsilon|$ if j is even and subtracting corresponding observations if j is odd.

It does not seem feasible to calculate general formulae for the method of grouping to give maximum efficiency. However, it is relatively simple to calculate the efficiency for a particular value of n and any chosen arrangement of groups, and hence to determine the best method of grouping for each n.

The important question is whether the efficiencies will be high enough to make the method a satisfactory substitute for the method of least squares. The maximum efficiency for large n (greater than about 50) is found to be practically constant for each coefficient. The efficiencies are listed below, and are seen to be all in the region of 90 per cent. For smaller n the efficiencies tend to be somewhat higher than these values.

a_0	100%	b_{20}	93.9%	b_{40}	91.4%
a_1	88.9%	b_{31}	92.5%	b_{51}	91.2%
a_2	89.7%	b_{42}	89.5%		
a_3	90.1%	b_{53}	91.2%		
a_4	90.4%				
a_5	90.6%				

The efficiency of the estimate $u_p(x)$ of the fitted value at a point varies somewhat with the location of the point, but is always close to 90%.

The coefficients b are linear functions of the coefficients a, but the method of grouping which gives the greatest efficiency for the coefficients a does not in general correspond to that method which would give maximum efficiency for a particular b_{pj}, since the coefficients a_j, unlike the corresponding least squares coefficients, are not orthogonal. But the choice of more complicated weights leads to only a slight improvement in the efficiency, and the method using weights w_j independent of p is much more convenient.

4. Tables and illustrative example. The following quantities are tabulated for values of n from 7 to 55 and for polynomials up to the fifth degree:

(a) The best method of grouping for the estimation of a_j, together with the weights $w_j(\epsilon)$.

(b) The divisor $\sum_{\epsilon} w_j(\epsilon)\epsilon^j$.

(c) The coefficients β_{kj}.

The coefficients β_{kj} are not in general integers. β_{20} and β_{31} are tabulated in full (r signifying that the last figure is repeated indefinitely), while β_{42} and β_{53} are given to ten significant figures and β_{40} and β_{51} to nine significant figures.

In the tables the observations are numbered by the value of $|\epsilon|$ if n is odd and by the value of $|\epsilon| + \frac{1}{2}$ if n is even. For example, for 62 observations the numbers are 1 to 31, for 63 observations 0 to 31. For even values of j observations of equal $|\epsilon|$ are added, while for odd values they are subtracted. This is indicated by the suffix + or − under the summation sign. The expression $c(a—b)$means that the observations numbered a to b (inclusive) are to be grouped and multiplied by the weight c.

It is convenient to illustrate the use of the tables by a specific example. We shall use the example of Birge and Shea [6], the measurements being the frequencies of the first 25 lines of the P branch of a CuH band. The frequencies vary from 22,330.52 to 23,295.47. After subtracting the constant amount 22,300 from each observation, the values are written down as in Model Form 1, starting from the bottom of the left-hand column and working up this column down the right-hand column.

The groups, weights, and β-coefficients are then entered in Model Form 2. Lines are drawn in Model Form 1 to indicate the sums required. For example, $\sum_{+}(6—10)$ occurs in a_4, so a line is drawn to the right between 5 and 6 and another line between 10 and 11.

The sums of the corresponding terms in the columns are added starting from the top, the progressive totals being entered at the right wherever a line is drawn. The differences between the corresponding terms are next added, the progressive totals being entered at the left. The required sums are obtained as differences of the progressive totals; for example,

$$\sum_{+}(6—10) = 12873.55 - 7017.03.$$

As a check, the right-hand and left-hand columns are added. If the sums are R and L, the final $\sum_{-}$ total should be $R - L$, the final $\sum_{+}$ total $R + L(+\, y(0))$.

The calculations indicated in Model Form 2 are then carried through. It is not necessary to record the actual sums. In working out a_2 on a calculating machine, the steps are

$$11(14968.38) - 11(11765.71) - 6(7017.03) \div 7370.$$

TABLE OF WEIGHTS

j = 0 | *j = 1*

$a_0 = \underset{+}{\Sigma} y/n$ $\qquad$ $a_1 = \underset{-}{\Sigma} w_1 y / \underset{-}{\Sigma} w_{1\epsilon}$

$b_{p0} = a_0 + \beta_{20} a_2 + \beta_{40} a_4$ $\qquad$ $b_{p1} = a_1 + \beta_{31} a_3 + \beta_{51} a_5$

n	β_{20}	β_{40}	w_1^*	$\underset{-}{\Sigma} w_{1\epsilon}$	β_{31}	β_{51}
7	−4	10.56	1(2—3)	10	−7	26.6666667
8	−5.25	21.8352273	1(2—4)	15	−8.25	46.5625
9	−6.6*r*	32.3478261	1(2—4)	18	−11	86.6666667
10	−8.25	56.0782895	1(3—5)	21	−14.25	152.0625
11	−10	56.4324324	1(2—5)	28	−16	171.578947
12	−11.916*r*	92.4430970	1(3—6)	32	−19.75	277.198864
13	−14	118.484210	1(3—6)	36	−24	400.307692
14	−16.25	178.9125	1(3—7)	45	−26.25	555.757415
15	−18.6*r*	259.2	1(3—7)	50	−31	761.176471
16	−21.25	363.085227	1(4—8)	55	−36.25	1054.98355
17	−24	440.228571	1(3—8)	66	−39	1200.25
18	−26.916*r*	591.596399	1(4—9)	72	−44.75	1550.67361
19	−30	596.689655	1(4—9)	78	−51	2053.08475
20	−33.25	700.141810	1(4—10)	91	−54.25	2471.49762
21	−36.6*r*	915.2	1(4—10)	98	−61	2677.97849
22	−40.25	1174.97540	1(5—11)	105	−68.25	3303.74133
23	−44	1485	1(4—11)	120	−72	4017.72973
24	−47.916*r*	1699.51705	1(5—12)	128	−79.75	4867.71991
25	−52	2103.90448	1(5—12)	136	−88	5462
26	−56.25	2575.15754	1(5—13)	153	−92.25	6518.06250
27	−60.6*r*	2593.69038	1(5—13)	162	−101	7717.36937
28	−65.25	2902.9725	1(6—14)	171	−110.25	9321.53708
29	−70	3505.46342	1(5—14)	190	−115	10591.5486
30	−74.916*r*	4194.79821	1(6—15)	200	−124.75	12598.6870
31	−80	4653.87610	1(6—15)	210	−135	13542.5807
32	−85.25	5500.39123	1(6—16)	231	−140.25	15553.9937
33	−90.6*r*	6455.21590	1(6—16)	242	−151	17882.4828
34	−96.25	7105.31250	1(7—17)	253	−162.25	20818.9764
35	−102	7099.2	1(6—17)	276	−168	23577
36	−107.916*r*	8260.38603	1(7—18)	288	−179.75	24189.5306
37	−114	9554.95384	1(7—18)	300	−192	27885.1777
38	−120.25	10396.6582	1(7—19)	325	−198.25	28898.9048
39	−126.6*r*	11927.1611	1(7—19)	338	−211	33115.0058
40	−133.25	13617.3606	1(8—20)	351	−224.25	37136.3294
41	−140	14739.4967	1(7—20)	378	−231	41372.4706
42	−146.916*r*	14734.2725	1(8—21)	392	−244.75	46135.2764
43	−154	16722.8852	1(8—21)	406	−259	52035.7419
44	−161.25	18901.3609	1(8—22)	435	−266.25	57467.5970
45	−168.6*r*	20294.6356	1(8—22)	450	−281	60364.8750
46	−176.25	22803.7383	1(9—23)	465	−296.25	67631.1107
47	−184	25533.8680	1(8—23)	496	−304	73189.0720
48	−191.916*r*	28497.1290	1(9—24)	512	−319.75	81530.2345
49	−200	30428.0980	1(9—24)	528	−336	89405.9520
50	−208.25	30449.5231	1(9—25)	561	−344.25	93132.9528
51	−216.6*r*	33843.3333	1(9—25)	578	−361	96082.0856
52	−225.25	37506.4392	1(10—26)	595	−378.25	104845.616
53	−234	39823.0675	1(9—26)	630	−387	114079.560
54	−242.916*r*	43950.4347	1(10—27)	648	−404.75	124075.662
55	−252	48384	1(10—27)	666	−423	136438.061

* $c(a—b)$ means that observations numbered a to b (inclusive) are to be grouped and multiplied by the weight c.

TABLE OF WEIGHTS—*Continued*

$j = 2$

$a_2 = \sum_{+} w_2 y / \sum_{+} w_2 \epsilon^2$ $b_{p2} = a_2 + \beta_{42} a_4$

n	w_2^*	$\sum_{+} w_2 \epsilon^2$	β_{42}
7	3(3) − 2(0—1)	50	−9.64
8	2(4) − 1(1—2)	44	−13.40909091
9	3(4) − 2(0—1)	92	−16.65217391
10	2(5) − 1(1—2)	76	−21.44736842
11	5(4—5) − 4(0—2)	370	−23.44324324
12	3(5—6) − 2(1—3)	268	−29.00746269
13	5(5—6) − 4(0—2)	570	−33.46315789
14	3(6—7) − 2(1—3)	400	−40.06
15	7(6—7) − 4(0—3)	1,078	−47.28571429
16	2(7—8) − 1(1—4)	352	−55.13636364
17	7(7—8) − 4(0—3)	1,470	−61.34285714
18	2(8—9) − 1(1—4)	472	−70.22881356
19	3(7—9) − 2(0—4)	1,044	−73.68965517
20	4(8—10) − 3(1—4)	1,624	−80.70689655
21	3(8—10) − 2(0—4)	1,350	−90.76
22	5(9—11) − 3(1—5)	2,480	−101.4419355
23	11(9—11) − 6(0—5)	5,984	−112.75
24	5(10—12) − 3(1—5)	3,080	−121.5181818
25	11(10—12) − 6(0—5)	7,370	−133.8597015
26	2(11—13) − 1(1—6)	1,452	−146.8305785
27	13(10—13) − 8(0—6)	12,428	−151.7531381
28	3(11—14) − 2(1—6)	3,200	−161.74
29	13(11—14) − 8(0—6)	14,924	−175.8780488
30	7(12—15) − 4(1—7)	8,624	−190.6428571
31	13(12—15) − 8(0—6)	17,628	−201.9734513
32	7(13—16) − 4(1—7)	10,136	−217.7707182
33	15(13—16) − 8(0—7)	23,140	−234.1972342
34	7(14—17) − 4(1—7)	11,760	−246.8714286
35	3(13—17) − 2(0—7)	6,250	−253
36	8(14—18) − 5(1—8)	17,680	−270.5941176
37	17(14—18) − 10(0—8)	39,780	−288.8153846
38	8(15—19) − 5(1—8)	20,240	−302.7086957
39	17(15—19) − 10(0—8)	45,390	−321.9617978
40	9(16—20) − 5(1—9)	25,320	−341.8440758
41	17(16—20) − 10(0—8)	51,340	−357.0821192
42	3(16—21) − 2(1—9)	10,800	−364.54
43	19(16—21) − 12(0—9)	71,858	−385.5901639
44	5(17—22) − 3(1—10)	19,840	−407.2677419
45	19(17—22) − 12(0—9)	80,522	−423.7239264
46	5(18—23) − 3(1—10)	22,180	−446.4329125
47	7(18—23) − 4(0—10)	32,466	−469.7710220
48	11(19—24) − 6(1—11)	53,284	−493.7369942
49	7(19—24) − 4(0—10)	35,994	−511.9404901
50	11(19—25) − 7(1—11)	65,604	−520.8661972
51	23(19—25) − 14(0—11)	142,968	−546
52	12(20—26) − 7(1—12)	77,672	−571.7602740
53	23(20—26) − 14(0—11)	157,458	−591.1840491
54	12(21—27) − 7(1—12)	85,400	−617.9780328
55	25(21—27) − 14(0—12)	184,800	−645.4

* $c(a—b)$ means that observations numbered a to b (inclusive) are to be grouped and multiplied by the weight c.

TABLE OF WEIGHTS—*Continued*

$j = 3$

$a_3 = \Sigma \underline{w}_3 y / \Sigma \underline{w}_3 \epsilon^3$ $\qquad$ $b_{p3} = a_3 + \beta_{35} a_5$

n	w_3^*	$\Sigma \underline{w}_3 \epsilon^3$	β_{35}
7	1(3) − 1(1—2)	36	−11.66666667
8	9(4) − 7(1—3)	504	−15.83333333
9	3(4) − 2(1—3)	240	−21
10	5(5) − 3(2—4)	540	−27.16666667
11	6(5) − 5(1—3)	1140	−30.47368421
12	15(6) − 11(2—4)	3630	−37.77272727
13	5(6) − 3(1—4)	1560	−44.84615385
14	24(7) − 13(2—5)	9204	−53.51694915
15	15(7) − 7(1—5)	7140	−61.94117647
16	7(8) − 3(2—6)	3990	−71.97368421
17	7(8) − 4(2—5)	5376	−78.75
18	35(9) − 17(2—6)	32130	−88.72222222
19	20(9) − 9(2—6)	21240	−100.8644068
20	48(10) − 19(2—7)	59736	−112.1946565
21	27(9—10) − 19(2—7)	63612	−117.6881720
22	63(10—11) − 40(2—8)	172620	−129.7846715
23	5(10—11) − 3(2—8)	15540	−144.1351351
24	20(11—12) − 11(2—9)	71280	−157.5740741
25	35(11—12) − 23(2—8)	154560	−167.25
26	77(12—13) − 48(3—9)	378840	−183.7195122
27	44(12—13) − 25(2—9)	244200	−198.7297297
28	24(13—14) − 13(3—10)	147264	−216.5677966
29	2(13—14) − 1(2—10)	13716	−232.9265092
30	117(14—15) − 56(3—11)	881244	−252.1282528
31	54(14—15) − 29(2—10)	485460	−262.2043011
32	39(15—16) − 20(3—11)	382590	−282.5244648
33	65(15—16) − 31(2—11)	701220	−301.4137931
34	35(16—17) − 16(3—12)	409920	−323.1065574
35	25(16—17) − 11(3—12)	316800	−345.7916667
36	5(16—18) − 3(3—13)	93060	−352.4432624
37	88(16—18) − 51(3—13)	1,768272	−375.9644670
38	11(17—19) − 7(3—13)	256410	−388.0855856
39	44(17—19) − 27(3—13)	1,102464	−412.7298851
40	64(18—20) − 37(3—14)	1,736928	−435.3016360
41	34(18—20) − 19(3—14)	988380	−461.3137255
42	17(19—21) − 9(3—15)	533052	−485.2363184
43	39(19—21) − 20(3—15)	1,305720	−512.6129032
44	247(20—22) − 123(4—16)	8,810490	−540.9827586
45	13(20—22) − 7(3—15)	524160	−555.5416667
46	247(21—23) − 129(4—16)	10,578516	−585.0301205
47	133(21—23) − 66(3—16)	6,091932	−611.8587896
48	56(22—24) − 27(4—17)	2,717820	−642.7178952
49	50(22—24) − 23(3—17)	2,587500	−670.8986667
50	280(23—25) − 141(4—17)	16,009140	−690.6405672
51	147(22—25) − 94(4—17)	10,971492	−703.9420655
52	105(23—26) − 64(4—18)	8,336160	−733.0989520
53	165(23—26) − 98(4—18)	13,809180	−766.9110070
54	44(24—27) − 25(4—19)	3,907200	−797.4189189
55	92(24—27) − 51(4—19)	8,595744	−832.5982533

* $c(a—b)$ means that observations numbered a to b (inclusive) are to be grouped and multiplied by the weight c.

TABLE OF WEIGHTS—*Continued*

$$j = 4$$
$$b_{p4} = a_4 = \sum_{+} w_4 y / \sum_{+} w_4 \epsilon^4$$

n	w_4^*	$\sum_{+} w_4 \epsilon^4$
7	2(3) − 5(2) + 2(0—1)	168
8	1(4) − 2(3) + 1(1)	144
9	25(4) − 46(3) + 14(0—1)	5376
10	9(5) − 10(3—4) + 11(1)	3600
11	71(5) − 73(3—4) + 50(0—1)	39648
12	16(6) − 15(4—5) + 14(1)	12480
13	72(6) − 53(3—5) + 58(0—1)	84768
14	59(7) − 41(4—6) + 32(1—2)	90000
15	45(7) − 29(4—6) + 28(0—1)	89880
16	20(8) − 11(4—7) + 12(1—2)	54960
17	59(8) − 31(4—7) + 26(0—2)	200376
18	144(9) − 71(5—8) + 70(1—2)	613152
19	590(9) − 395(4—7) + 396(0—2)	4,138824
20	206(10) − 131(5—8) + 106(1—3)	1,721280
21	415(10) − 245(5—8) + 226(0—2)	4,182864
22	284(11) − 161(6—9) + 120(1—3)	3,345552
23	235(11) − 117(5—9) + 100(0—3)	3,395784
24	415(12) − 194(6—10) + 185(1—3)	7,072128
25	217(12) − 98(6—10) + 78(0—3)	4,241328
26	288(13) − 115(6—11) + 134(1—3)	6,855936
27	2989(13) − 1155(6—11) + 1126(0—3)	80,879904
28	16(14) − 6(7—12) + 5(1—4)	489312
29	2989(14) − 1344(6—11) + 1450(0—3)	125,116488
30	944(15) − 410(7—12) + 379(1—4)	44,279712
31	535(15) − 221(7—12) + 226(0—3)	28,384776
32	586(16) − 235(8—13) + 206(1—4)	34,551408
33	1533(16) − 561(7—13) + 532(0—4)	103,700520
34	[illegible]3(16—17) − 502(8—14) + 497(1—4)	97,162464
35	1491(16—17) − 957(8—14) + 826(0—4)	209,629728
36	2240(17—18) − 1405(9—15) + 1071(1—5)	346,514448
37	9420(17—18) − 5397(8—15) + 5408(0—4)	1680,885360
38	141(18—19) − 79(9—16) + 70(1—5)	27,560016
39	11148(18—19) − 6045(9—16) + 5792(0—4)	2408,213808
40	2820(19—20) − 1765(9—16) + 1696(1—5)	763,126848
41	13332(19—20) − 8151(9—16) + 7008(0—5)	3922,631856
42	415(20—21) − 245(10—17) + 226(1—5)	133,851648
43	15620(20—21) − 9031(10—17) + 7456(0—5)	5452,591056
44	4779(21—22) − 2576(10—18) + 2271(1—6)	1855,459872
45	19074(21—22) − 9955(10—18) + 9354(0—5)	8071,065288
46	5535(22—23) − 2834(11—19) + 2406(1—6)	2521,801584
47	21945(22—23) − 10923(11—19) + 9894(0—5)	10844,513064
48	415(23—24) − 194(11—20) + 185(1—6)	226,308096
49	8489(23—24) − 4667(11—19) + 3850(0—6)	5386,084704
50	3785(24—25) − 1693(12—21) + 1560(1—6)	2388,695712
51	30485(24—25) − 15249(11—20) + 14080(0—6)	22841,591136
52	8680(25—26) − 4263(12—21) + 3610(1—7)	6940,591392
53	34645(25—26) − 16549(12—21) + 14800(0—6)	29764,941336
54	1967(26—27) − 924(13—22) + 758(1—7)	1798,408080
55	46255(26—27) − 20515(12—22) + 17754(0—7)	46002,560208

* $c(a—b)$ means that observations numbered a to b (inclusive) are to be grouped and multiplied by the weight c.

TABLE OF WEIGHTS—*Continued*

$$j = 5$$
$$b_{55} = a_5 = \Sigma w_5 y / \Sigma w_{5e}^2$$

n	w_5^*	Σw_{5e}^2
7	1(3) − 4(2) + 5(1)	240
8	15(4) − 49(3) + 35(1—2)	6720
9	9(4) − 26(3) + 14(1—2)	6720
10	49(5) − 111(4) + 84(1—2)	65520
11	26(5) − 55(4) + 30(1—2)	51840
12	72(6) − 143(5) + 55(1—3)	208560
13	3(6) − 4(4—5) + 3(1—3)	15120
14	112(7) − 130(5—6) + 143(2—3)	840840
15	275(7) − 301(5—6) + 231(1—3)	2,808960
16	19(8) − 20(6—7) + 13(2—4)	252720
17	481(8) − 464(6—7) + 364(1—3)	8,910720
18	1755(9) − 1360(6—8) + 1547(2—4)	47,895120
19	287(9) − 213(6—8) + 189(1—4)	9,954000
20	819(10) − 589(7—9) + 456(2—5)	35,112000
21	112(10) − 75(7—9) + 68(1—4)	6,283200
22	697(11) − 455(8—10) + 357(2—5)	47,295360
23	1037(11) − 583(7—10) + 561(1—5)	95,729040
24	5586(12) − 3059(8—11) + 2622(2—6)	615,593160
25	510(12) − 430(8—10) + 366(1—5)	92,085120
26	228(13) − 115(9—12) + 100(2—6)	37,044000
27	1824(13) − 1274(8—11) + 1235(2—6)	485,503200
28	2528(14) − 1701(9—12) + 1413(2—7)	788,492880
29	3220(14) − 2030(9—12) + 2009(2—6)	1202,742240
30	35200(15) − 21518(10—13) + 18183(2—7)	15207,265920
31	22149(15) − 13230(10—13) + 10235(2—7)	10916,650080
32	14553(16) − 8463(11—14) + 5735(2—8)	8211,661920
33	29475(16) − 16344(11—14) + 12800(2—7)	19440,662400
34	34125(17) − 16632(11—15) + 15125(2—8)	27440,028000
35	10465(17) − 4998(11—15) + 4199(2—8)	9460,956960
36	3540(18) − 1652(12—16) + 1239(2—9)	3608,463600
37	1925(18) − 861(12—16) + 732(2—8)	2257,995600
38	36736(19) − 14874(12—17) + 14245(2—9)	51234,261120
39	110374(19) − 43890(12—17) + 39121(2—9)	171065,664000
40	12111(20) − 4719(13—18) + 3809(2—10)	20909,168280
41	69223(20) − 25960(13—18) + 23405(2—9)	135787,454880
42	524160(21) − 241736(13—18) + 229395(3—10)	1,379383,716480
43	17949(21) − 8085(13—18) + 6944(2—10)	52229,469600
44	18954(22) − 8385(14—19) + 6794(3—11)	60473,424960
45	6075(22) − 2568(14—19) + 2233(2—10)	21840,477120
46	104091(23) − 43290(15—20) + 35445(3—11)	408520,153080
47	8060(23) − 3289(15—20) + 2461(2—11)	34633,959360
48	128(24) − 47(15—21) + 47(3—11)	639,576000
49	73437(24) − 26468(15—21) + 24192(2—11)	400419,714240
50	34595(25) − 12285(16—22) + 10619(3—12)	204535,553040
51	9709(25) − 3400(16—22) + 2793(3—12)	62103,392400
52	57967(25—26) − 36800(17—23) + 30355(3—12)	696524,337600
53	13650(25—26) − 8075(16—23) + 7514(3—12)	185253,868800
54	289960(26—27) − 168883(17—24) + 144768(3—13)	4,267843,419840
55	411312(26—27) − 236698(17—24) + 193397(3—13)	6,521985,180480

* $c(a—b)$ means that observations numbered a to b (inclusive) are to be grouped and multiplied by the weight c.

MODEL FORM 1

	$\underset{-}{\Sigma}$				$\underset{+}{\Sigma}$
0			647.29		
1	*81.67*	605.48		687.15	
2		561.83		725.15	
3		516.42		761.27	*4504.59*
4	*816.01*	469.22		795.39	
5		420.29		827.54	*7017.03*
6		369.60		857.71	
7		317.17		885.85	
8	*2928.93*	263.06		911.94	
9		207.40		935.95	*11765.71*
10	*4465.36*	149.98		957.86	*12873.55*
11		91.05		977.79	*13942.39*
12	*6317.05*	30.52		995.47	*14968.38*
		4002.02		*10319.07*	

MODEL FORM 2*

b_0	$\underset{+}{\Sigma}(0—12)/25$	a_0	598.735200
	β_{20} $-$ 52	$\beta_{20}a_2$	$+48.492011$
	β_{40} $+$ 2103.90448	$\beta_{40}a_4$	$+0.029212$
		b_0	647.256423
b_2	$\underset{+}{\Sigma}\{11(10—12) - 6(0—5)\}/7370$	a_2	-0.9325387
	β_{42} $-$ 133.8597015	$\beta_{42}a_4$	-0.0018586
		b_2	-0.9343973
b_4	$\underset{+}{\Sigma}\{217(12—\quad) - 98(6—10) + 78(0—3)\}/4{,}241328$		
		$a_4 = b_4$	0.0000138848
b_1	$\underset{-}{\Sigma}1(5—12)/136$	a_1	40.448824
	β_{31} $-$ 88	$\beta_{31}a_3$	$+0.385928$
	β_{51} $+$	$\beta_{51}a_5$	
		b_1	40.834752
b_3	$\underset{-}{\Sigma}\{35(11—12) - 23(2—8)\}/154560$	a_3	-0.004385546
	β_{53} $-$	$\beta_{53}a_5$	
		b_3	-0.004385546
b_5	$\underset{-}{\Sigma}\{(\quad—\quad) - (\quad—\quad) + (\quad—\quad)\}/$		
		$a_5 = b_5$	

* $c(a—b)$ means that observations numbered a to b (inclusive) are to be grouped and multiplied by the weight c.

If a calculating machine is not available, it is best to work out the sums $(a—b)$ individually. The product $w(a—b)$ should be multiplied out in full, but

seven-figure logarithms may be used for the division by $\Sigma w_j \epsilon^j$ and for the calculation of the terms $\beta_{kj} a_k$.

The values obtained for the polynomial coefficients by the method of least squares and by the method of grouping are shown below, together with the standard errors.

	Least Squares	*Grouping*
$b_0 \times 10^2$	647254.3 ± 12	647256.4
$b_1 \times 10^3$	40834.6 ± 2.2	40834.8
$b_2 \times 10^4$	-9343.5 ± 5	-9344.0
$b_3 \times 10^6$	-437.6 ± 2.2	-438.6
$b_4 \times 10^8$	13.8 ± 3.6	13.9

When the polynomial coefficients have been determined, the fitted values may be worked out. If the polynomial is required in terms of a variable other than ϵ, a Horner shift is performed in the usual way.

If the standard errors are required, the residuals v_p must be calculated. Assuming an efficiency of 90%, the estimated standard error of an observation is given by

$$s_p = [\Sigma v_p^2 / \{n - 0.9(p + 1)\}]^{\frac{1}{2}}.$$

The estimated errors of the polynomial coefficients and fitted values can be found by using the tabulated weight functions [7] for the least squares solution, multiplied by the factor 1.05 to allow for the efficiency of 90%.

It is sometimes necessary to know whether the neglect of higher powers is justified. The quantities $a_j'^2 \Sigma T_j^2$ provide a test for determining the degree of the polynomial to be used, since $a_j'^2 \Sigma T_j^2$ is the amount by which Σv^2 is reduced when the degree is increased from $j - 1$ to j in the least squares method. To a sufficiently good approximation we can use a_j for a_j' and put $\Sigma T_j^2 = n^{2j+1}/\kappa_j$, where

$$\kappa_1 = 12, \quad \kappa_2 = 180, \quad \kappa_3 = 2800, \quad \kappa_4 = 44{,}000, \quad \kappa_5 = 700{,}000.$$

In the example used here we find that

$$a_3^2 \Sigma T_3^2 = 41; \qquad a_4^2 \Sigma T_4^2 = 0.016; \qquad a_5^2 \Sigma T_5^2 = 0.009.$$

Thus a_3 is highly significant, a_4 is of doubtful significance, while terms of higher degree are probably insignificant.

REFERENCES

[1] R. A. Fisher and F. Yates, *Statistical Tables for Biological, Agricultural, and Medical Research*, 3rd ed., Oliver and Boyd, Ltd., Edinburgh, 1948.

[2] R. L. Anderson and E. E. Houseman, "Tables of orthogonal polynomial values extended to $n = 104$," *Research Bulletin 297* (1942), Iowa State College Agricultural Experimental Station.

[3] R. T. Birge, "Least-squares fitting of data by means of polynomials," *Rev. Modern Physics*, Vol. 19 (1947), pp. 298–360.

[4] Harold Jeffreys, *Theory of Probability*, 2nd ed., Clarendon Press, Oxford, 1948, p. 179.
[5] P. G. Guest, "Orthogonal polynomials in the least-squares fitting of observations," *Philos. Mag.* (7), Vol. 41 (1950), pp. 124–137.
[6] R. T. Birge and J. D. Shea, "A rapid method for calculating the least squares solution of a polynomial of any degree," *Univ. California Publ. Math.*, Vol. 2 (1927), pp. 67–118.
[7] P. G. Guest, "Estimation of the error at a point on a least-squares curve," *Australian Jour. Sci. Research. Ser. A*, Vol. 3 (1950), pp. 173–182; "Estimation of the errors of the least-squares polynomial coefficients," *ibid*, pp. 364–375.

A MULTIVARIATE GAMMA-TYPE DISTRIBUTION[1]

By A. S. Krishnamoorthy and M. Parthasarathy

Madras Christian College, Tambaram, India, and
Ramanujan Institute of Mathematics, Madras, India

Introduction. Mehler has shown that the two-variate probability density function (pdf) for correlated variates, each of which has a marginal Guassian distribution, can be expressed as a series bilinear in Hermite polynomials:

$$\frac{1}{2\pi\sqrt{1-\rho^2}} \exp\{-\tfrac{1}{2}(x^2 - 2\rho xy + y^2)/(1-\rho^2)\}$$
$$= \frac{1}{2\pi} \exp\{-\tfrac{1}{2}(x^2 + y^2)\}\left[1 + \rho H_1(x)H_1(y) + \frac{\rho^2}{2!} H_2(x)H_2(y) + \cdots\right].$$

Kibble [5] has extended this result to any number of variables and noticed a small difference between the generalization and the particular case due to Mehler.

It is known that Mehler's series is not an isolated result, there being a similar series bilinear in Laguerre polynomials, discussed by Hardy [3], Watson [6], and Kibble [4], and series bilinear in certain other other polynomials, discussed by Campbell [2], and by Aitken and Gonin [1]. All these results can be generalized for any number of variables in much the same way as Kibble has generalized Mehler's result. These generalizations are contained in Krishnamoorthy's thesis "Multivariate Distribution Functions" (in the library of the University of Madras). In the present paper the generalization involving Laguerre polynomials is given.

1. Notation and summary. It was shown by Kibble [4] that a two-variate distribution function, in which each of the variates x_i, $i = 1, 2$, has the frequncy function

$$\phi(x_i) = \frac{x_i^{p-1} e^{-x_i}}{\Gamma(p)}, \tag{1.1}$$

may be represented by

$$\phi(x_1)\phi(x_2)\left[1 + \frac{\rho^2}{p} L_1(x_1, p)L_1(x_2, p) + \frac{\rho^4}{2!p(p+1)} L_2(x_1, p)L_2(x_2, p) + \cdots\right],$$

where $L_r(x, p)$, $p > 0$, is the generalized Laguerre polynomial of degree r satisfying

$$L_r(x, p) \equiv r!L_r^{(p-1)}(x) = \frac{\left(-\frac{d}{dx}\right)^r [x^r\phi(x)]}{\phi(x)}. \tag{1.2}$$

[1] Sections 1 to 4 of this paper, deriving the distributions, were written by the first author; Section 5, on the convergence of certain series, was contributed by the second author.

It is the object of this paper to extend Kibble's result to n variables, assuming (i) that the variates have each a marginal Gamma-type distribution given by (1.1) with the same parameter p; (ii) that the variates have Gamma-type distributions with different parameters. The extension in case (i) appears in (3.7) and that in case (ii) appears in (4.1). The convergence of series obtained in either case is established in Section 5.

An outline of the procedure followed may be given thus. We obtain, in (2.2), the moment-generating function (mgf) for the joint distribution of $\xi_i = \frac{1}{2}x_i^2$, $(i = 1, 2, \cdots, n)$, where each x_i has a normal distribution with zero mean. From this we get the mgf for the distribution of the sums of squares in a sample of m from a normal correlated n-variate distribution, and thence, in (2.3), a possible mgf for an n-variate distribution in which each variate has a Gamma-type distribution. Finally we obtain from (2.3) the n-variate distribution in (3.7) by a process which is essentially the inversion of the Laplace transform,

$$(1.3) \qquad (1-\alpha)^{-p}\left(\frac{\alpha}{1-\alpha}\right)^r = \int_0^\infty e^{\alpha x} f_r(x, p)\, dx,$$

where

$$f_r(x, p) = \frac{L_r(x, p)\phi(x)}{p^{(r)}}, \qquad p^{(r)} = p(p+1)\cdots(p+r-1).$$

It will be noticed that (1.3) is true for $r = 0$ if we define (as we may) $L_0(x, p) = 1$, $p^0 = 1$.

2. An mgf for an n-variate Gamma-type distribution. Let $\| \rho_{ij} \|$ defined by

$$\| \rho_{ij} \| = \begin{Vmatrix} 1 & \rho_{12} & \cdots & \rho_{1n} \\ \rho_{21} & \rho_{22} & \cdots & \rho_{2n} \\ \cdot & \cdot & & \\ \cdot & \cdot & & \\ \cdot & \cdot & & \\ \rho_{n1} & \rho_{n2} & \cdots & \rho_{nn} \end{Vmatrix},$$

where $\rho_{ij} = \rho_{ji}$, be a positive definite matrix. Then the normal correlated n-variate distribution having zero means and $\| \rho_{ij} \|$ for its variance-covariance matrix is given by

$$(2.1) \qquad dF = \frac{| \rho^{ij} |^{\frac{1}{2}}}{(2\pi)^{\frac{1}{2}n}} \exp\left\{-\tfrac{1}{2} \sum_{i,j=1}^{n} \rho^{ij} x_i x_j\right\} dx_1\, dx_2 \cdots dx_n$$

in the usual notation, where $\| \rho^{ij} \|$ is the inverse of $\| \rho_{ij} \|$. Denoting the mgf for a distribution of ξ_i, $i = 1, 2, \cdots, n$, having any pdf, by

$$G(\alpha_1, \alpha_2, \cdots, \alpha_n) = E\,\{\exp \sum_i \alpha_i \xi_i\},$$

we find that, in the case $\xi_i = \frac{1}{2}x_i^2$ where x_i have the joint distribution (2.1), the mgf is

$$G_{\frac{1}{2}}(\alpha_1, \alpha_2, \cdots, \alpha_n) = E\{\exp \tfrac{1}{2} \sum \alpha_i x_i^2\}$$

$$= \frac{|\rho^{ij}|^{\frac{1}{2}}}{(2\pi)^{\frac{1}{2}n}} \int \exp\{-\tfrac{1}{2} \sum (\rho^{ij} - \delta_{ij}\alpha_i)x_i x_j\}\, d\bar{x},$$

where $-\infty \leq x_i \leq \infty$, $i = 1, 2, \cdots, n$, $\int \cdots d\bar{x}$ denotes an integration with respect to all x_i, and δ_{ij} is the Kronecker delta defined as zero if $i \neq j$ and unity otherwise. Therefore

$$G_{\frac{1}{2}}(\alpha_1, \alpha_2, \cdots, \alpha_n) = |\rho^{ij}|^{\frac{1}{2}} |\rho^{ij} - \delta_{ij}\alpha_i|^{-\frac{1}{2}} = |\delta_{ij} - \rho_{ij}\alpha_i|^{-\frac{1}{2}}, \tag{2.2}$$

provided that $\| \rho^{ij} - \delta_{ij}\alpha_i \|$ is positive definite. It follows that the mgf for the distribution of the sums of squares in a sample of m from a normal correlated n-variate distribution is obtained by raising the expression in (2.2) to the mth power, and furthermore, that the replacement of $m/2$ by p leads to a possible mgf for an n-variate distribution in which each variate x_i has the frequency function (1.1). Therefore *a possible mgf for an n-variate Gamma-type distribution defined as above is obtained from* (2.2) *when* $-\frac{1}{2}$ *in the power on the right side of* (2.2) *is changed to* $-p$. *The expression for the mgf is*

$$G_p(\alpha_1, \alpha_2, \cdots, \alpha_n) = |\delta_{ij} - \rho_{ij}\alpha_i|^{-p}$$

$$= \begin{vmatrix} 1-\alpha_1 & -\rho_{12}\alpha_2 & \cdots & -\rho_{1n}\alpha_n \\ -\rho_{12}\alpha_1 & 1-\alpha_2 & \cdots & -\rho_{2n}\alpha_n \\ \cdot & \cdot & & \\ \cdot & \cdot & & \\ \cdot & \cdot & & \\ -\rho_{1n}\alpha_1 & -\rho_{2n}\alpha_2 & \cdots & 1-\alpha_n \end{vmatrix}^{-p} \tag{2.3}$$

$$= \{(1-\alpha_1)(1-\alpha_2)\cdots(1-\alpha_n)\}^{-p} \begin{vmatrix} 1 & -\rho_{12}\beta_2 & \cdots & -\rho_{1n}\beta_n \\ -\rho_{12}\beta_1 & 1 & \cdots & -\rho_{2n}\beta_n \\ \cdot & \cdot & & \\ \cdot & \cdot & & \\ \cdot & \cdot & & \\ -\rho_{1n}\beta_1 & -\rho_{2n}\beta_2 & \cdots & 1 \end{vmatrix}^{-p},$$

where $\beta_i = \alpha_i/(1 - \alpha_i)$, $i = 1, 2, \cdots, n$. It is convenient to write (2.3) in the form

$$\begin{aligned} &G(\alpha_1, \alpha_2, \cdots, \alpha_n) \\ &\quad = \{(1-\alpha_1)(1-\alpha_2)\cdots(1-\alpha_n)\}^{-p}\{g(\beta_1, \beta_2, \cdots, \beta_n)\}^{-p}, \end{aligned} \tag{2.4}$$

where $g(\beta_1, \beta_2, \cdots, \beta_n)$ is the determinant of β's in (2.3).

3. A series for an n-variate Gamma-type distribution. Expanding the g in (2.4) by Maclaurin's theorem for a function of n variables, we get

$$g(\beta_1, \beta_2, \cdots, \beta_n) = g_0 + \left(\sum_i \beta_i \frac{\partial}{\partial \beta_i}\right) g_0 \tag{3.1}$$
$$+ \frac{1}{2!}\left(\sum_i \beta_i \frac{\partial}{\partial \beta_i}\right)^{[2]} g_0 + \cdots + \frac{1}{n!}\left(\sum_i \beta_i \frac{\partial}{\partial \beta_i}\right)^{[n]} g_0,$$

where $g_0 = g(0, 0, \cdots, 0)$ and $(\Sigma_i \beta_i(\partial/\partial\beta_i))^{[r]} g_0$ is the result of first expanding $(\Sigma \beta_i(\partial/\partial\beta_i))^r$ regarding the operators $\partial/\partial\beta_i$ as algebraical numbers, then giving the operators their proper roles in the expanded form of $(\Sigma \beta_i(\partial/\partial\beta_i))^r$, and finally putting $\beta_1 = \beta_2 = \cdots = \beta_n = 0$ in the partial derivatives of g which we get when the expanded form is applied to g.

Clearly the expansion of $g(\beta_1, \beta_2, \cdots, \beta_n)$ in (3.1) does not contain terms linear in the β's or terms such as $\beta_1^{p_1} \cdots \beta_i^{p_i} \cdots$ with any $p_i > 1$. In fact

$$g_0 = 1, \qquad \frac{\partial g_0}{\partial \beta_1} = \begin{vmatrix} 0 & 0 & \cdots & 0 \\ -\rho_{12} & 1 & \cdots & 0 \\ \cdot & \cdot & & \\ \cdot & \cdot & & \\ \cdot & \cdot & & \\ -\rho_{1n} & 0 & \cdots & 1 \end{vmatrix} = 0,$$

where, of course, the partial derivation is performed before we put $\beta_1 = \beta_2 = \cdots = \beta_n = 0$; and similarly $\partial g_0/\partial\beta_2$, $\partial g_0/\partial\beta_3$, $\cdots$, $\partial g_0/\partial\beta_n$ are all zero, so that $(\Sigma_{i=1}^n \partial/\partial\beta_i) g_0 = 0$.
Further

$$\frac{\partial^2 g_0}{\partial\beta_1 \partial\beta_2} = \begin{vmatrix} 0 & -\rho_{12} & 0 & \cdots & 0 \\ -\rho_{12} & 0 & 0 & \cdots & 0 \\ \cdot & \cdot & & & \\ \cdot & \cdot & & & \\ \cdot & \cdot & & & \\ -\rho_{1n} & -\rho_{2n} & & & 1 \end{vmatrix} = \begin{vmatrix} 0 & \rho_{12} \\ \rho_{12} & 0 \end{vmatrix} = -C_{12} \quad \text{say},$$

$$\frac{\partial^3 g_0}{\partial\beta_1 \partial\beta_2 \partial\beta_3} = -\begin{vmatrix} 0 & \rho_{12} & \rho_{13} \\ \rho_{12} & 0 & \rho_{23} \\ \rho_{13} & \rho_{23} & 0 \end{vmatrix} = -C_{123} \quad \text{say},$$

$$\cdot$$
$$\cdot$$
$$\cdot$$

$$\frac{\partial^n g_0}{\partial\beta_1 \partial\beta_2 \cdots \partial\beta_n} = (-1)^n \begin{vmatrix} 0 & \rho_{12} & \cdots & \rho_{1n} \\ \rho_{12} & 0 & & \rho_{2n} \\ \cdot & \cdot & & \\ \cdot & \cdot & & \\ \cdot & \cdot & & \\ \rho_{1n} & \rho_{2n} & \cdots & 0 \end{vmatrix} = -C_{123\cdots n} \quad \text{say}.$$

Hence (3.1) can be written

$$g(\beta_1, \beta_2, \cdots, \beta_n) = 1 - \Big(\sum_{i<j} C_{ij}\, \beta_i\, \beta_j + \sum_{i<j<k} C_{ijk}\, \beta_i\, \beta_j\, \beta_k \tag{3.2}$$
$$+ \cdots + C_{12\cdots n}\, \beta_1\, \beta_2 \cdots \beta_n = 1 - B \text{ say}.$$

Using (3.2) in (2.4) and expanding $(1 - B)^{-p}$ formally by the binomial theorem, we get

$$G_p(\alpha_1, \alpha_2, \cdots, \alpha_n) = \{(1 - \alpha_1)(1 - \alpha_2) \cdots (1 - \alpha_n)\}^{-p} \tag{3.3}$$
$$\cdot \left\{ \sum_0^\infty \frac{p(p+1) \cdots (p+r-1)}{r!} B^r \right\} = \sum_0^\infty \frac{p(p+1) \cdots (p+r-1)}{r!} B_r^*,$$

where

$$B_r^* = \{(1 - \alpha_1)(1 - \alpha_2) \cdots (1 - \alpha_n)\}^{-p} B^r.$$

Expanding B^r by the multinomial theorem, we can express B_r^* as a finite sum of the form

$$B_r^* = \prod_{i=1}^n (1 - \alpha_i)^{-p} \left(\sum K \beta_1^{a_1} \beta_1^{a_2} \cdots \beta_n^{a_n}\right) \tag{3.4}$$
$$= \sum K \left\{ \prod_{i=1}^n (1 - \alpha_i)^{-p} \left(\frac{\alpha_i}{1 - \alpha_i} \right)^{-a_i} \right\}$$

where K is a polynomial in $C_{ij}, \cdots, C_{12\cdots n}$, and $a_1, \cdots, a_n$ are nonnegative integers of which not more than $n - 2$ are zero.

It is now plain from (3.4) and (1.3) that B_r^* can be expressed as a Laplace transform:

$$B_r^* = \int_{0 \le x_i \le \infty} e^{\Sigma \alpha_i x_i} \sum K \prod_{i=1}^n f_{a_i}(x_i, p)\, d\bar{x},$$

of which the determining function is

$$\sum K \prod_{i=1}^n f_{a_i}(x_i, p) = \phi(x_1)\phi(x_2) \cdots \phi(x_n) \sum K \prod_{i=1}^n \frac{L_{a_i}(x_i, p)}{p^{(a_i)}} \tag{3.5}$$

$$= \phi(x_1)\phi(x_2) \cdots \phi(x_n) \left\{ \sum_{i<j} C_{ij} \frac{L(x_i, p)}{p} \frac{L(x_j, p)}{p} \right. \tag{3.6}$$
$$\left. + \cdots + C_{12\cdots n} \frac{L(x_1, p)}{p} \cdots \frac{L(x_n, p)}{p} \right\}^r,$$

where $\{\cdots\}^r$ is a symbol for the rth power of a multinomial, in expanding which we suppose that

$$\left\{\frac{L(x, p)}{p}\right\}^m \left\{\frac{L(x, p)}{p}\right\}^n = \left\{\frac{L(x, p)}{p}\right\}^{m+n}$$

for all positive integers m, n, and after expanding which we set

$$\left\{\frac{L(x, p)}{p}\right\}^m \equiv \frac{L_m(x, p)}{p^{(m)}}.$$

Finally we can replace B_r^* in the series of (3.3) by its determining function in (3.6) and obtain the form

$$\phi(x_1)\phi(x_2) \cdots \phi(x_n) \sum_0^\infty \frac{p^{(r)}}{r!}\left\{\sum_{i<j} C_{ij} \frac{L(x_i, p)}{p} \frac{L(x_j, p)}{p} + \cdots + C_{123\cdots n} \frac{L(x_i, p)}{p} \cdots \frac{L(x_n, p)}{p}\right\}^r, \tag{3.7}$$

where $\phi(x_i)$ is defined by (1.1), for a distribution function having $G_p(\alpha_1, \alpha_2, \cdots, \alpha_n)$ defined as in (2.3) for its mgf. The convergence of series (3.7) is proved, with a certain restriction on the ρ's, in Section 5. Consequently, *with this restriction as regards convergence, we can take* (3.7) *to be an n-variate distribution function in which each variate* x_i, $i = 1, 2, \cdots, n$ *has the distribution function* $\phi(x_i)$ *in* (1.1).

Remark on the series (3.7). If there are only two β's present in any term of (3.4), this being their least number possible, they will be raised to the same degree r, and therefore the corresponding term of (3.5) will have Laguerre polynomials of the same degree r. If, however, more than two β's are present in a term of (3.4), their degrees may be different and consequently also the degrees of the Laguerre polynomials in the corresponding term of (3.5). Hence the n-variate Gamma-type distribution symbolically denoted by (3.7) has the property that (i) any term in its expansion involving two variables contains Laguerre polynomials of the same degree in those variables, while (ii) a term involving more than two variables may contain Laguerre polynomials of different degrees in the variables. It is known [5] that an analogous property is possessed by the extension to n variates of Mehler's series in Hermite polynomials.

4. A generalization of Section 3. If we take instead of the mgf in (2.4) the more general mgf

$$(1 - \alpha_1)^{-p_1}(1 - \alpha_2)^{-p_2} \cdots (1 - \alpha_n)^{-p_n}\{g(\beta_1, \beta_2, \cdots, \beta_n)\}^{-p}$$

and repeat the reasoning of Section 3, we shall obtain, in the symbolic notation of (3.7), the following series (whose convergence is established in Section 5 under the condition on the ρ's already referred to):

$$\phi(x_1)\phi(x_2) \cdots \phi(x_n) \sum_0^\infty \frac{p^{(r)}}{r!}\left\{\sum_{i<j} C_{ij} \frac{L(x_i, p_i)}{p_i} \frac{L(x_j, p_j)}{p_j} + \cdots + C_{12\cdots n} \frac{L(x_1, p_1)}{p_1} \cdots \frac{L(x_n, p_n)}{p_n}\right\}^r, \tag{4.1}$$

$$\phi(x_i) = \frac{x_i^{p_i-1}e^{-x_i}}{\Gamma(p_i)}, \qquad i = 1, 2, \cdots, n. \tag{4.2}$$

This series, under the condition which secures its convergence, may be regarded as an n-variate distribution function in which each variate x_i, $i = 1, 2, \cdots, n$ has the distribution function $\phi(x_i)$ of (4.2).[2]

5. Addendum: the convergence of the series in (3.7) and (4.1). The object of this addendum is to establish, under a suitable condition, the convergence of the series in (3.7) and (4.1) The proof of the convergence depends on the following lemma.

Lemma. *In the symbolic notation of* (3.6), *for $r \geq 1$*

$$\left|\left\{\frac{L(x, p)}{p}\right\}^r\right| \equiv \left|\frac{L_r(x, p)}{p^{(r)}}\right| < \begin{cases} K(x, p)r^{\frac{1}{2}(-p+\frac{1}{2})}, & 0 < p < \frac{1}{2} \\ K(x, p), & p \geq \frac{1}{2}, \end{cases}$$

where $K(x, p)$ is a constant depending on x and p.

Proof. From the well known result $\Gamma(x + a)/\Gamma(x) \sim x^a$ as $x \to \infty$, where a is a constant, we get

$$\frac{p^{(r)}}{r!} = \frac{p(p + 1) \cdots (p + r - 1)}{r!} = \frac{\Gamma(p + r)}{\Gamma(p)\Gamma(r + 1)} \sim \frac{r^{p-1}}{\Gamma(p)} \quad \text{as } r \to \infty. \tag{5.1}$$

From a formula of Fejér [7], Hille [8] has deduced that

$$\frac{L_r(x, p)}{r!} = \frac{1}{\sqrt{\pi}} e^{\frac{1}{2}x} x^{-\frac{1}{2}(p-\frac{1}{2})} r^{\frac{1}{2}(p-\frac{3}{2})} \cos\left[2\sqrt{rx} - \pi\left(\frac{1}{4} + \frac{p - 1}{2}\right)\right] + O\,[r^{\frac{1}{2}(p-2)}], \quad r \to \infty. \tag{5.2}$$

Combining (5.2) with (5.1), we conclude that

$$\left|\frac{L_r(x, p)}{p^{(r)}}\right| = \left|\frac{L_r(x, p)/r!}{p^{(r)}/r!}\right| < A(x, p)r^{-\frac{1}{2}(p-\frac{1}{2})}, \qquad r > r_0, \tag{5.3}$$

where $A(x, p)$ is a constant which depends on x and p. Further, once r_0 is fixed,

$$\left|\frac{L_r(x, p)}{p^{(r)}}\right| < B(x, p), \qquad r \leq r_0, \tag{5.4}$$

where $B(x, p)$ is also a constant which depends on x and p. Equations (5.3) and (5.4) together yield the result stated in the lemma where $K = \max(A, B)$.

Theorem. *The series in* (3.7),

$$\sum \frac{p^{(r)}}{r!} t_r \equiv \sum \frac{p^{(r)}}{r!}\left\{\sum C_{ij} \frac{L(x_i, p)}{p} \frac{L(x_j, p)}{p} + \cdots + C_{123\cdots n} \frac{L(x_1, p)}{p} \cdots \frac{L(x_n, p)}{p}\right\}^r,$$

is absolutely convergent provided that

---

[2] Thanks are due to Dr. P. Kesava Menon and Prof. C. T. Rajagopal for helping to settle certain points of detail.

(5.5) $$\sigma \equiv \sum_{i,j} | C_{ij}| + \sum_{i,j,k} | C_{ijk} | + \cdots + | C_{123\cdots n} | < 1.$$

Proof. We have, in symbolic notation,

(5.6) $$t_r = \sum \lambda_{m_2,m_2',\cdots,m_n} \left\{C_{ij} \frac{L(x_i, p)}{p} \frac{L(x_j, p)}{p}\right\}^{m_2} \left\{C_{i'j'} \frac{L(x_i', p)}{p} \frac{L(x_j', p)}{p}\right\}^{m_2'} \cdots \left\{C_{123\cdots n} \frac{L(x_1, p)}{p} \cdots \frac{L(x_n, p)}{p}\right\}^{m_n},$$

where one at least of the suffixes i', j' is different from i, j (similar statements being true of the C's with 3, 4, $\cdots$ suffixes), and

$$\lambda_{m_2,m_2',\cdots m_n} = \frac{r!}{m_2!\, m_2'! \cdots m_n!}, \qquad m_2 + m_2' + \cdots + m_n = r.$$

First suppose that $p \geq \frac{1}{2}$. Then (5.6) gives, by virtue of the lemma,

(5.7) $$| t_r | \leq \Sigma\lambda_{m_2,m_2',\cdots,m_n} K(x_1, p)K(x_2, p) \cdots K(x_n, p) \, | C_{ij} |^{m_2} \, | C_{i'j'} |^{m_2'} \cdots | C_{123\cdots n} |^{m_n}.$$

Therefore, writing

$$\kappa = \max \{K(x_1, p), K(x_2, p), \cdots, K(x_n, p)\}$$

we get from (5.7)

$$| t_r | \leq \kappa^n \Sigma\lambda_{m_2, m_2',\cdots,m_n} | C_{ij} |^{m_2} | C_{i'j'} |^{m_2'} \cdots | C_{123\cdots n} |^{m_n}$$
$$= \kappa^n(\Sigma \, | C_{ij} | + \Sigma \, | C_{ijk} | + \cdots + | C_{123\cdots n} |)^r \equiv \kappa^n\sigma^r.$$

And so $(p^{(r)}/r!) \, | t_r | \leq u_r \equiv (p^{(r)}/r!)\kappa^n\sigma^r$, where Σu_r is known to be convergent for $\sigma < 1$, and hence $\Sigma p^{(r)} t_r/r!$ is absolutely convergent for $\sigma < 1$.

In the case $p < \frac{1}{2}$, it is obvious from the lemma that

$$\frac{p^{(r)}}{r!} \, | t_r | \leq v_r \equiv \frac{p^{(r)}}{r!} \kappa^n r^{\frac{1}{2}(\frac{1}{2}-p)n} \sigma^r,$$

where

$$v_r^{1/r} = \left[\frac{p^{(r)}}{r!}\right]^{1/r} \kappa^{n/r} \left[r^{1/r}\right]^{\frac{1}{2}(\frac{1}{2}-p)n} \sigma \to \sigma \text{ as } r \to \infty.$$

Consequently, by Cauchy's root-test, Σv_r is convergent for $\sigma < 1$, and so again $\Sigma p^{(r)} t_r/r!$ is absolutely convergent for $\sigma < 1$.

A sufficient condition for the convergence of the series in the theorem, simpler in form than (5.5), *is*

(5.8) $$N\rho^* < 1,$$

where N is the result of replacing every one of the ρ's in the C's by unity and ρ^* is the maximum of the terms in the ρ's when we omit the numerical coefficients of the terms.

A sufficient condition for the convergence of the series (4.1) *is again either* (5.5) *or* (5.8) since, arguing exactly as above, we find that

$$| \text{ the } (r+1)^{th} \text{ term of the series (4.1) } | \leq \prod_{i=1}^{n} \phi(x_i).\ \frac{p^{(r)}}{r!}\, \kappa^n r^{\Sigma_i \frac{1}{2}(\frac{1}{2}-p_i)} \sigma^r,$$

where the summation in the power of r is for all p_i which are less than $\frac{1}{2}$.

Note. The case $n = 2$ makes the series in the theorem identical with a series obtained by W. F. Kibble [4] for a two-variate Gamma-type distribution. Kibble's proof of the convergence is, however defective[3] since he assumes that

$$\frac{L_r(x, p)}{p^{(r)}} \sim \frac{L_{r-1}(x, p)}{p^{(r-1)}}, \qquad r \to \infty,$$

is a consequence of (5.2).

REFERENCES

[1] A. C. Aitken and H. T. Gonin, "On fourfold sampling with or without replacement," *Proc. Roy. Soc. Edinburgh. Sect. A*, Vol. 55 (1935), pp. 114–125.

[2] J. T. Campbell, "The Poisson correlation function," *Proc. Edinburgh Math. Soc. Series 2*, Vol. 4 (1934), pp. 18–26.

[3] G. H. Hardy, "Summation of series of polynomials of Laguerre," *Jour. London Math. Soc.*, Vol. 7 (1932), pp. 138–140.

[4] W. F. Kibble, "A two-variate Gamma-type distribution," *Sankhyā*, Vol. 5 (1941), pp. 137–150.

[5] W. F. Kibble, "An extension of a theorem of Mehler on Hermite polynomials," *Proc. Cambridge Philos. Soc.*, Vol. 41 (1945), pp. 12–15.

[6] G. N. Watson, "Notes on the generating functions of polynomials: (1) Laguerre polynomials," *Jour. London Math. Soc.*, Vol. 8 (1933), pp. 189–192.

[7] L. Fejér. "Sur une méthode de M. Darboux," *C. R. Acad. Sci. Paris*, Vol. 147 (1908), pp. 1040–1042.

[8] E. Hille, "On Laguerre's series. First note," *Proc. Nat. Acad. Sci.*, Vol. 12 (1926), pp. 261–265.

[3] Acknowledgement is due to Prof. C. T. Rajagopal for having drawn attention to this defect and suggested a method of removing it.

A COMBINATORIAL CENTRAL LIMIT THEOREM[1]

By Wassily Hoeffding

Institute of Statistics, University of North Carolina

1. Summary. Let $(Y_{n1}, \cdots, Y_{nn})$ be a random vector which takes on the $n!$ permutations of $(1, \cdots, n)$ with equal probabilities. Let $c_n(i, j)$, $i, j = 1, \cdots, n$, be n^2 real numbers. Sufficient conditions for the asymptotic normality of

$$S_n = \sum_{i=1}^{n} c_n(i, Y_{ni})$$

are given (Theorem 3). For the special case $c_n(i, j) = a_n(i)b_n(j)$ a stronger version of a theorem of Wald, Wolfowitz and Noether is obtained (Theorem 4). A condition of Noether is simplified (Theorem 1).

2. Introduction and statement of results. An example of what is here meant by a combinatorial central limit theorem is a solution of the following problem. For every positive integer n there are given $2n$ real numbers $a_n(i)$, $b_n(i)$, $i = 1, \cdots, n$. It is assumed that the $a_n(i)$ are not all equal and the $b_n(i)$ are not all equal. Let $(Y_{n1}, \cdots, Y_{nn})$ be a random vector which takes on the $n!$ permutations of $(1, \cdots, n)$ with equal probabilities $1/n!$. Under what conditions is

$$S_n = \sum_{i=1}^{n} a_n(i)b_n(Y_{ni}) \tag{1}$$

asymptotically normally distributed as $n \to \infty$?

Throughout this paper a random variable S_n will be called asymptotically normal or asymptotically normally distributed if

$$\lim_{n\to\infty} \Pr\{S_n - ES_n \leq x\sqrt{\operatorname{var} S_n}\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}y^2}\, dy, \qquad -\infty < x < \infty,$$

where ES_n and var S_n are the mean and the variance of S_n.

In the particular case $a_n(i) = b_n(i) = i$ the asymptotic normality of S_n was proved by Hotelling and Pabst [2]. The first general result is due to Wald and Wolfowitz [6], who showed that S_n is asymptotically normal if, as $n \to \infty$,

$$\frac{\frac{1}{n}\sum_{i=1}^{n}(a_n(i) - \bar{a}_n)^r}{\left[\frac{1}{n}\sum_{i=1}^{n}(a_n(i) - \bar{a}_n)^2\right]^{r/2}} = O(1), \qquad r = 3, 4, \cdots, \tag{2}$$

and

$$\frac{\frac{1}{n}\sum_{i=1}^{n}(b_n(i) - \bar{b}_n)^r}{\left[\frac{1}{n}\sum_{i=1}^{n}(b_n(i) - \bar{b}_n)^2\right]^{r/2}} = O(1), \qquad r = 3, 4, \cdots, \tag{3}$$

[1] Work done under the sponsorship of the Office of Naval Research.

where

$$\bar{a}_n = \frac{1}{n} \sum_{i=1}^{n} a_n(i), \qquad \bar{b}_n = \frac{1}{n} \sum_{i=1}^{n} b_n(i).$$

Noether [5] proved that condition (3) can be replaced by the weaker condition

$$\lim_{n\to\infty} \frac{\sum_{i=1}^{n} (b_n(i) - \bar{b}_n)^r}{\left[\sum_{i=1}^{n} (b_n(i) - \bar{b}_n)^2\right]^{r/2}} = 0, \qquad r = 3, 4, \cdots . \tag{4}$$

This condition can be simplified as follows.

THEOREM 1. *Condition* (4) *is equivalent to either of the following two conditions:*

$$\lim_{n\to\infty} \frac{\sum_{i=1}^{n} | b_n(i) - \bar{b}_n |^r}{\left[\sum_{i=1}^{n} (b_n(i) - \bar{b}_n)^2\right]^{r/2}} = 0 \qquad \textit{for some } r > 2; \tag{5}$$

$$\lim_{n\to\infty} \frac{\max_{1 \le i \le n} (b_n(i) - \bar{b}_n)^2}{\sum_{i=1}^{n} (b_n(i) - \bar{b}_n)^2} = 0. \tag{6}$$

Hence conditions (2) *and* (5) *or* (2) *and* (6) *are sufficient for the asymptotic normality of* (1).

The proof is given in Section 3. For a more general condition and a stronger but simpler condition see Theorem 4 below.

One extension of this problem was considered by Daniels [1], who studied the asymptotic distribution of

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_n(i, j) b_n(Y_{ni}, Y_{nj}).$$

The present paper is concerned with an alternative extension. It considers the distribution of

$$S_n = \sum_{i=1}^{n} c_n(i, Y_{ni}), \tag{7}$$

where $c_n(i, j)$, $i, j = 1, \cdots, n$, are n^2 real numbers, defined for every positive integer n. In the particular case $c_n(i, j) = a_n(i)b_n(j)$, (7) reduces to (1).

Let

$$d_n(i, j) = c_n(i, j) - \frac{1}{n} \sum_{g=1}^{n} c_n(g, j) - \frac{1}{n} \sum_{h=1}^{n} c_n(i, h) + \frac{1}{n^2} \sum_{g=1}^{n} \sum_{h=1}^{n} c_n(g, h). \tag{8}$$

THEOREM 2. *The mean and variance of*

$$S_n = \sum_{i=1}^{n} c_n(i, Y_{ni})$$

are

$$(9)\qquad ES_n = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} c_n(i, j),$$

$$(10)\qquad \operatorname{var} S_n = \frac{1}{n-1}\sum_{i=1}^{n}\sum_{j=1}^{n} d_n^2(i, j).$$

Henceforth we assume that $d_n(i, j) \neq 0$ for some (i, j), so that var $S_n > 0$. In the special case $c_n(i, j) = a_n(i)b_n(j)$ this corresponds to the assumption that the $a_n(i)$ are not all equal and the $b_n(j)$ are not all equal.

THEOREM 3. *The distribution of* $S_n = \sum_{i=1}^n c_n(i, Y_{ni})$ *is asymptotically normal if*

$$(11)\qquad \lim_{n\to\infty} \frac{\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} d_n^r(i, j)}{\left[\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} d_n^2(i, j)\right]^{r/2}} = 0, \qquad r = 3, 4, \cdots.$$

Condition (11) *is satisfied if*

$$(12)\qquad \lim_{n\to\infty} \frac{\max_{1\le i,j\le n} d_n^2(i, j)}{\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} d_n^2(i, j)} = 0.$$

Theorems 2 and 3 will be proved in Sections 4 and 5.

For the special case $c_n(i, j) = a_n(i)b_n(j)$, Theorem 3 immediately gives

THEOREM 4. *The distribution of* $S_n = \sum_{i=1}^n a_n(i)b_n(Y_{ni})$ *is asymptotically normal if*

$$(13)\qquad \lim_{n\to\infty} n^{\frac{1}{2}r-1} \frac{\sum_{i=1}^{n}(a_n(i) - \bar{a}_n)^r}{\left[\sum_{i=1}^{n}(a_n(i) - \bar{a}_n)^2\right]^{r/2}} \frac{\sum_{i=1}^{n}(b_n(i) - \bar{b}_n)^r}{\left[\sum_{i=1}^{n}(b_n(i) - \bar{b}_n)^2\right]^{r/2}} = 0, \quad r = 3, 4, \cdots.$$

Condition (13) *is satisfied if*

$$(14)\qquad \lim_{n\to\infty} n \frac{\max_{1\le i\le n}(a_n(i) - \bar{a}_n)^2}{\sum_{i=1}^{n}(a_n(i) - \bar{a}_n)^2} \frac{\max_{1\le i\le n}(b_n(i) - \bar{b}_n)^2}{\sum_{i=1}^{n}(b_n(i) - \bar{b}_n)^2} = 0.$$

It will be observed that the symmetrical condition (13) contains Noether's condition (2) and (4) as a special case.

Let $X_n = (X_{n1}, \cdots, X_{nn})$ be independent of and have the same distribution as $Y_n = (Y_{n1}, \cdots, Y_{nn})$.

THEOREM 5. *The random variable*

$$(15)\qquad S_n' = \sum_{i=1}^{n} c_n(X_{ni}, Y_{ni})$$

has the same distribution as S_n *in* (7).

In fact, the conditional distribution of S'_n given that $X_n = p$, a fixed permutation of $(1, \cdots, n)$, is independent of p because the distribution of Y_n is invariant under permutations of its components.

The distribution of sums of the form (1) has attracted the attention of statisticians in connection with nonparametric tests (see, for example, [2], [6], [3]) and sampling from a finite population (which leads to the case $a_n(i) = 0$ for $i > m$; cf. also Madow [4]). More general sums of the form (7) or (15) are likewise of interest in nonparametric theory. Thus it follows from results of Lehmann and Stein [3] that a test of the hypothesis that $U_1, \cdots, U_n$ are independent and identically distributed, which is most powerful similar against the alternative that the joint frequency function is $f_1(u_1) \cdots f_n(u_n)$ is based on a statistic of the form (7) with

$$c_n(i, j) = \log f_i(u_j),$$

where the u_j are the observed sample values. If the n pairs $(U_1, V_1), \cdots, (U_n, V_n)$ are independent and identically distributed, a test of the hypothesis that U_i and V_i are independent which is most powerful similar against the alternative that their joint frequency function is $f(u, v)$ is based on a statistic of the form (15) with $c_n(i, j) = \log f(u_i, v_j)$, where $(u_1, v_1), \cdots, (u_n, v_n)$ are the observed values.

In these examples the numbers $c_n(i, j)$ are random variables. An application of some of the present results to such cases will be considered by the author in a forthcoming paper.

3. Proof of Theorem 1. Let

$$g_i = \frac{b_n(i) - \bar{b}_n}{\left[\sum_{i=1}^{n} (b_n(i) - \bar{b}_n)^2\right]^{1/2}},$$

$$G_n = \max(|g_1|, \cdots, |g_n|).$$

Theorem 1 asserts the equivalence of the three relations

$$\lim_{n\to\infty} \sum_{i=1}^{n} g_i^r = 0, \qquad r = 3, 4, \cdots; \tag{16}$$

$$\lim_{n\to\infty} \sum_{i=1}^{n} |g_i|^r = 0 \qquad \text{for some } r > 2; \tag{17}$$

$$\lim_{n\to\infty} G_n = 0. \tag{18}$$

We have

$$\sum_{i=1}^{n} g_i^2 = 1,$$

and hence for $r > 2$

$$G_n^r \le \sum_{i=1}^{n} |g_i|^r \le G_n^{r-2} \sum_{i=1}^{n} g_i^2 = G_n^{r-2}.$$

The equivalence of (16), (17) and (18) follows immediately.

4. Proof of Theorem 2. The subscript n in Y_{ni}, $c_n(i, j)$, etc., will henceforth be omitted. We note that if the subscripts $i_1, \cdots, i_m$ are distinct, the expected value of a function $f(Y_{i_1}, \cdots, Y_{i_m})$ is equal to

$$\frac{1}{n(n-1)\cdots(n-m+1)} \sum_{j_1,\cdots,j_m}' f(j_1, \cdots, j_m),$$

where the sum Σ' is extended over all m-tuples $(j_1, \cdots, j_m)$ of distinct integers from 1 to n. Relation (9) follows immediately.

Let

$$T_n = \sum_{i=1}^{n} d(i, Y_i), \tag{19}$$

where $d(i, j) = d_n(i, j)$ is defined by (8). Using (9), we get

$$T_n = S_n - ES_n. \tag{20}$$

Also

$$\sum_{i=1}^{n} d(i, j) = 0 \text{ for all } j, \qquad \sum_{j=1}^{n} d(i, j) = 0 \text{ for all } i. \tag{21}$$

Hence

$$Ed(i, Y_i) = 0,$$

$$Ed^2(i, Y_i) = \frac{1}{n} \sum_{j=1}^{n} d^2(i, j),$$

and if $i \neq j$,

$$\begin{aligned} Ed(i, Y_i)d(j, Y_j) &= \frac{1}{n(n-1)} \sum_{g,h}' d(i, g)d(j, h) \\ &= \frac{-1}{n(n-1)} \sum_{g=1}^{n} d(i, g)d(j, g). \end{aligned}$$

Therefore

$$\begin{aligned} \operatorname{var} S_n = \operatorname{var} T_n &= \sum_{i=1}^{n} Ed^2(i, Y_i) + \sum_{i,j}' Ed(i, Y_i)d(j, Y_j) \\ &= \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} d^2(i, j) - \frac{1}{n(n-1)} \sum_{g=1}^{n} \sum_{i,j}' d(i, g)d(j, g) \\ &= \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} d^2(i, j) + \frac{1}{n(n-1)} \sum_{g=1}^{n} \sum_{i=1}^{n} d^2(i, g), \end{aligned}$$

which gives relation (10).

5. Proof of Theorem 3. Let

$$M_{r,n} = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} d^r(i, j), \tag{22}$$

$$\bar{M}_{r,n} = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} | d(i, j) |^r, \tag{23}$$

$$D_n = \max_{1 \le i,j \le n} | d(i, j) |. \tag{24}$$

Then var $S_n = n/(n-1)\, M_{2,n}$. Since, by hypothesis, var $S_n > 0$, we may and shall assume that

$$M_{2,n} = 1. \tag{25}$$

Conditions (11) and (12) can now be written as

$$\lim_{n\to\infty} M_{r,n} = 0, \qquad r = 3, 4, \cdots, \tag{26}$$

and

$$\lim_{n\to\infty} D_n = 0. \tag{27}$$

That (27) implies (26) is seen from the inequalities

$$| M_{r,n} | \le \bar{M}_{r,n} \le D_n^{r-2} M_{2,n} = D_n^{r-2} \qquad \text{for } r > 2.$$

Since

$$\bar{M}^2_{2k+1,n} \le M_{2k,n} M_{2k+2,n}, \qquad k = 1, 2, \cdots,$$

condition (26) implies

$$\lim_{n\to\infty} \bar{M}_{r,n} = 0, \qquad r = 3, 4, \cdots. \tag{28}$$

As var $S_n \to 1$, it is now sufficient to demonstrate that under conditions (25) and (28), $T_n = S_n - ES_n$ has a normal limiting distribution with mean 0 and variance 1. This will be proved by showing that

$$\lim_{n\to\infty} ET_n^r = \begin{cases} 1\cdot 3 \cdots (r-1) & \text{if } r \text{ is even,} \\ 0 & \text{if } r \text{ is odd.} \end{cases} \tag{29}$$

The rth moment of T_n,

$$ET_n^r = E \sum_{i_1=1}^{n} \cdots \sum_{i_r=1}^{n} d(i_1, Y_{i_1}) \cdots d(i_r, Y_{i_r}), \tag{30}$$

can be written as a sum of terms of the form

$$I(r, e_1, \cdots, e_m) = \sum_{i_1,\cdots,i_m}{}' E d^{e_1}(i_1, Y_{i_1}) \cdots d^{e_m}(i_m, Y_{i_m}), \tag{31}$$

where $e_i \ge 1$, $e_1 + \cdots + e_m = r$. The number of terms (31) is independent of n. It will be shown that

$$\lim_{n\to\infty} I(r, e_1, \cdots, e_m) = 0 \qquad \text{unless } r = 2m, \qquad e_1 = \cdots = e_m = 2, \tag{32}$$

$$\lim_{n\to\infty} I(r, 2, \cdots, 2) = 1 \qquad \text{if } r \text{ even,} \tag{33}$$

and that the number of terms $I(r, 2, \cdots, 2)$ in (30) with r even equals $1 \cdot 3 \cdots (r-1)$. Then (29) holds, and the theorem will be proved.

We have for $n \to \infty$

$$I(r, e_1, \cdots, e_m) \sim n^{-m} \sum_{i_1, \cdots, i_m}' \sum_{j_1, \cdots, j_m}' d^{e_1}(i_1, j_1) \cdots d^{e_m}(i_m, j_m). \tag{34}$$

The right-hand side can be written as a sum of terms which, apart from the sign, are of the form

$$\begin{aligned} &n^{-m} J(r, p, q, e_1, \cdots, e_m) \\ &\quad = n^{-m} \sum_{i_1=1}^{n} \cdots \sum_{i_p=1}^{n} \sum_{j_1=1}^{n} \cdots \sum_{j_q=1}^{n} d^{e_1}(i_{c_1}, j_{d_1}) \cdots d^{e_m}(i_{c_m}, j_{d_m}), \end{aligned} \tag{35}$$

where

$$1 \le p \le m, \qquad 1 \le q \le m,$$

$$1 \le c_g \le p, \qquad 1 \le d_h \le q, \qquad (g, h = 1, \cdots, m),$$

and for every integer u, $1 \le u \le p (1 \le u \le q)$ at least one $c_g(d_h)$ is equal to u. The number of terms (35) is independent of n.

The sum J in (35) can be written as a product of $s \ge 1$ sums of a similar form,

$$J(r, p, q, e_1, \cdots, e_m) = \prod_{k=1}^{s} J(r_k, p_k, q_k, e_{k1}, \cdots, e_{km_k}), \tag{36}$$

where

$$(e_{k1}, \cdots, e_{km_k}), \qquad k = 1, \cdots, s,$$

are s disjoint subsets of $(e_1, \cdots, e_m)$,

$$\begin{gathered} e_{k1} + \cdots + e_{km_k} = r_k, \qquad r_1 + \cdots + r_s = r, \\ p_1 + \cdots + p_s = p, \qquad q_1 + \cdots + q_s = q, \\ m_1 + \cdots + m_s = m. \end{gathered} \tag{37}$$

We observe that

$$1 \le p_k \le m_k, \qquad 1 \le q_k \le m_k, \qquad m_k \le r_k. \tag{38}$$

It will be assumed that s is the greatest possible number of factors into which $J(r, p, q, e_1, \cdots, e_m)$ can be decomposed in the form (36). If $s = 1$, the number of equalities between the subscripts c or between the subscripts d in (35) must be at least $m - 1$. The total number of subscripts c, d being $2m$, there are at most $m + 1$ distinct subscripts, so that $p + q \le m + 1$. If

$$(c_g, d_g) = (c_h, d_h) \qquad \text{for some } (g, h), g \ne h, \tag{39}$$

we have strict inequality. For an arbitrary s we have in a similar way

$$p_k + q_k \le m_k + 1, \qquad k = 1, \cdots, s, \tag{40}$$

and hence

$$(41) \qquad p + q \leq m + s,$$

with strict inequality in the case (39).

By Hölder's inequality, from (35),

$$| J(r, p, q, e_1, \cdots, e_m) | \leq \prod_{g=1}^{m} \Big(\sum_{i_1} \cdots \sum_{i_p} \sum_{j_1} \cdots \sum_{s} | d(i_{c_g}, j_{d_g}) |^r\Big)^{e_g/r}$$
$$= \prod_{g=1}^{m} (n^{p+q-1} \bar{M}_{r,n})^{e_g/r} = n^{p+q-1} \bar{M}_{r,n}.$$

Similarly,

$$| J(r_k, p_k, q_k, e_{k1}, \cdots, e_{km_k}) | \leq n^{p_k+q_k-1} \bar{M}_{r_k,n}.$$

Hence, by (36),

$$(42) \qquad n^{-m} | J(r, p, q, e_1, \cdots, e_m) | \leq n^{p+q-s-m} \bar{M}_{r_1,n} \cdots \bar{M}_{r_s,n}.$$

If, for some k, $r_k = 1$, then, by (38) and (37), $p_k = q_k = m_k = e_{k1} = 1$, and hence $J = 0$ by (21). Thus we may assume $r_k \geq 2$, $k = 1, \cdots, s$. Then, by (28), $\bar{M}_{r_1 n} \cdots \bar{M}_{r_s n} \to 0$ unless $r_1 = \cdots = r_s = 2$. It now follows from (42) and (41) that

$$(43) \qquad \lim_{n \to \infty} n^{-m} J(r, p, q, e_1, \cdots, e_m) = 0$$

except perhaps when $r_1 = \cdots = r_s = 2$.

If $r_1 = \cdots = r_s = 2$, we have

$$(44) \qquad n^{-m} J(r, p, q, e_1, \cdots, e_m) = O(n^{p+q-s-m}).$$

By (38), $r_k = 2$ implies $m_k = 1$ or 2. If $m_k = 2$, then $e_{k1} = e_{k2} = 1$ and $p_k + q_k \leq 3$ by (40). If $p_k + q_k = 3$, the corresponding J-factor is of the form

$$\sum_i \sum_j \sum_k d(i, j) d(i, k) \qquad \text{or} \qquad \sum_i \sum_j \sum_k d(i, k) d(j, k),$$

both of which vanish by (21). If $m_k = 2$ and $p_k + q_k = 2$, we have case (39) and hence, by the remark following (41), $p + q - s - m < 0$. By (44), this implies (43).

Thus the only case where (43) need not hold is $r_k = 2$, $m_k = 1$ for $k = 1, \cdots, s$. Then $p_k = q_k = 1$, $e_{k1} = 2$, hence

$$r = 2s = 2m, \qquad p = q = r/2$$
$$e_1 = \cdots = e_m = 2.$$

This proves relation (32), and (33) follows from

$$I(r, 2, \cdots, 2) \sim n^{-r/2} J\left(r, \frac{r}{2}, \frac{r}{2}, 2, \cdots, 2\right)$$
$$= n^{-r/2} [J(2, 1, 1, 2)]^{r/2}$$
$$= M_{2,n}^{r/2} = 1.$$

It remains to determine the number of terms $I(r, 2, \cdots, 2)$ in (30) when r is even. This is the number of ways the subscripts $i_1, \cdots, i_r$ can be tied in $r/2$ groups of two, which is $(r-1)(r-3)\cdots 3\cdot 1$. The proof is complete.

REFERENCES

[1] H. E. DANIELS, "The relation between measures of correlation in the universe of sample permutations," *Biometrika*, Vol. 33 (1944), pp. 129–135.

[2] H. HOTELLING AND M. PABST, "Rank correlation and tests of significance involving no assumption of normality," *Annals of Math. Stat.*, Vol. 7 (1936), pp. 29–43.

[3] E. L. LEHMANN AND C. STEIN, "On the theory of some nonparametric hypotheses," *Annals of Math. Stat.*, Vol. 20 (1949), pp. 28–45.

[4] W. G. MADOW, "On the limiting distributions of estimates based on samples from finite universes," *Annals of Math. Stat.*, Vol. 19 (1948), pp. 535–545.

[5] G. E. NOETHER, "On a theorem by Wald and Wolfowitz," *Annals of Math. Stat.*, Vol. 20 (1949), pp. 455–458.

[6] A. WALD AND J. WOLFOWITZ, "Statistical tests based on permutations of the observations," *Annals of Math. Stat.*, Vol. 15 (1944), pp. 358–372.

ON RATIOS OF CERTAIN ALGEBRAIC FORMS

By Robert V. Hogg

State University of Iowa

1. Introduction. In an investigation of the ratio of the mean square successive difference to the mean square difference in random samples from a normal universe with mean zero, J. D. Williams [4] proved the rather surprising fact that any moment of this ratio is equal to the corresponding moment of the numerator divided by that of the denominator. Later Tjallings Koopmans [2] and John von Neumann [3] showed independently that this ratio and its denominator are stochastically independent. From this, Williams' theorem is an immediate consequence. In this paper, we determine a necessary and sufficient condition for the stochastic independence of a ratio and its denominator. We then use this condition in our study of certain ratios of algebraic forms.

2. Stochastic independence of a ratio and its denominator. We prove the following theorem for the continuous type distribution. Consider two one-dimensional random variables x and y and their probability density function $g(x,y)$. Let $P(y \leqq 0) = 0$. Assume the moment generating function, $M(u,t) = E[\exp(ux + ty)]$, exists for $-T < u,t < T$, $T > 0$. The theorem is as follows.

Theorem 1. *Under the conditions stated, in order that y and $r = x/y$ be stochastically independent, it is necessary and sufficient that*

$$\frac{\partial^k M(0,t)}{\partial u^k} \equiv \frac{\dfrac{\partial^k M(0,0)}{\partial u^k}}{\dfrac{\partial^k M(0,0)}{\partial t^k}} \frac{\partial^k M(0,t)}{\partial t^k},$$

for $k = 0, 1, 2, \cdots$.

Proof of necessity. If $f(r, y)$ is the probability density function of the variables r and y, it is well known that a necessary and sufficient condition for the independence of the random variables r and y is that $f(r,y) \equiv f_1(r)f_2(y)$, where $f_1(r)$ and $f_2(y)$ are the marginal density functions of r and y respectively. Hence, since $x = ry$,

$$M(u,t) \equiv E[\exp(ury + ty)];$$

or

$$M(u,t) \equiv \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp(ury + ty) f_1(r) f_2(y)\, dr\, dy.$$

By hypothesis, the moments of x of order k exist; so

$$\frac{\partial^k M(0,t)}{\partial u^k} \equiv \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (ry)^k \exp(ty) f_1(r) f_2(y)\, dr\, dy.$$

Finally,

$$\frac{\partial^k M(0, t)}{\partial u^k} \equiv \int_{-\infty}^{\infty} r^k f_1(r) dr \cdot \int_{-\infty}^{\infty} y^k \exp(ty) f_2(y) dy,$$

for $k = 0, 1, 2, \cdots$. If we set $t = 0$, we see that $\int_{-\infty}^{\infty} r^k f_1(r) dr$ exists, since it is equal to the quotient of the kth moments of x and y,

$$K_k = \frac{\dfrac{\partial^k M(0, 0)}{\partial u^k}}{\dfrac{\partial^k M(0, 0)}{\partial t^k}}.$$

The hypothesis precludes the moments of y being zero. We also note that

$$\int_{-\infty}^{\infty} y^k \exp(ty) f_2(y) dy \equiv \frac{\partial^k M(0, t)}{\partial t^k};$$

consequently

$$\frac{\partial^k M(0, t)}{\partial u^k} \equiv K_k \frac{\partial^k M(0, t)}{\partial t^k},$$

for $k = 0, 1, 2, \cdots$.

PROOF OF SUFFICIENCY. Consider the identity

$$\frac{\partial^k M(0, t)}{\partial u^k} \equiv K_k \frac{\partial^k M(0, t)}{\partial t^k},$$

or

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^k \exp(ty) g(x,y) dx\, dy \equiv K_k \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y^k \exp(ty) g(x,y) dx\, dy. \tag{2.1}$$

Since all the moments of x and y exist, we may differentiate p times with respect to t under the integral signs. Then if we set $t = 0$,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^k y^p g(x,y) dx\, dy = K_k \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y^{k+p} g(x,y) dx\, dy,$$

for $p = 0, 1, 2, \cdots$. Although t has been restricted to the range $-T < t < T$, we may extend that range to $-\infty < t < T$ and still have the existence of $M(u,t)$. The condition that $P(y \leqq 0) = 0$ further permits us to integrate (2.1) $p' \leqq k$ times under the integral signs as shown below.

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{0}\int_{-\infty}^{t_{p'}} \cdots \int_{-\infty}^{t_2} x^k \exp(t_1 y) g(x, y) \prod_{j=1}^{p'} dt_j\, dx\, dy$$

$$= K_k \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{0}\int_{-\infty}^{t_{p'}} \cdots \int_{-\infty}^{t_2} y^k \exp(t_1 y) g(x,y) \prod_{j=1}^{p'} dt_j\, dx\, dy,$$

or

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(\frac{x}{y}\right)^{p'} \cdot x^{k-p'} g(x,y) dx\, dy = K_k \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y^{k-p'} g(x,y) dx\, dy,$$

for $p' = 1, 2, 3, \cdots, k$. These two expressions may be written

$$E(x^k y^m) = K_k E(y^{k+m})$$

for $k = 0, 1, 2, \cdots$ and $m = -k, \cdots, -1, 0, 1, 2, \cdots$. If $m = -k$, then

$$E\left[\left(\frac{x}{y}\right)^k\right] = K_k.$$

Thus

$$E(x^k y^m) = E\left[\left(\frac{x}{y}\right)^k\right] E(y^{k+m}),$$

or

$$E\left[\left(\frac{x}{y}\right)^k y^{k+m}\right] = E\left[\left(\frac{x}{y}\right)^k\right] \cdot E(y^{k+m}),$$

for $k = 0, 1, 2, \cdots$ and $m = -k, \cdots, -1, 0, 1, 2, \cdots$. This could also be rewritten as

$$E(r^k y^h) = E(r^k) \cdot E(y^h),$$

for $k = 0, 1, 2, \cdots$ and $h = 0, 1, 2, \cdots$. This is sufficient to insure stochastic independence of r and y; thus the proof is complete.

3. Ratios of linear forms in gamma variables. Let the independent random variables x_j have the gamma density functions

$$f_j(x_j) = \begin{cases} \dfrac{1}{\Gamma(c_j + 1) d_j^{c_j+1}} (x_j)^{c_j} \exp\left(-\dfrac{x_j}{d_j}\right), & 0 \leqq x < \infty, \\ 0, & \text{elsewhere}, \end{cases}$$

where $c_j > -1$ and $d_j > 0$, for $j = 1, 2, \cdots, n$. Construct the two real linear forms $L_1 = \sum_1^n a_j x_j$ and $L_2 = \sum_1^n b_j x_j$, $b_j > 0$. Let L_1 and L_2 be linearly independent; thus their ratio will not be a mere constant.

THEOREM 2. *Under the conditions stated, a necessary and sufficient condition that L_2 and L_1/L_2 be stochastically independent is that*

$$b_1 d_1 = b_2 d_2 = \cdots = b_n d_n .$$

PROOF. Our proof consists in showing, by the use of Theorem 1, that if some of the bd values are distinct, the variance of L_1/L_2 is equal to zero. This fact further implies that the ratio is a constant, and hence the necessity of the condition is proved by contradiction. For the sufficiency, we demonstrate that the partial derivatives of the moment generating function $E[\exp(uL_1 + tL_2)]$ satisfy the condition of Theorem 1. However in interest of conservation of paper, a referee[1] has suggested that upon setting $u_j^2 = x_j/d_j$, von Neumann's argument [3] may be made to complete the proof.

[1] We take this opportunity to thank the Referee for this and other suggestions.

An interesting consequence of Theorem 2 is the following corollary. Let $Q_1 = X'AX$ and $Q_2 = X'BX$ be two real symmetric quadratic forms in n random values of a variable normally distributed with mean zero. We restrict Q_2 to be nonnegative (or nonpositive). Let $AB = BA$. It is known ([1], p. 25) that there then exists an orthogonal matrix C such that simultaneously $C'AC$ and $C'BC$ are diagonal matrices formed by the characteristic numbers a_j of A and b_j of B respectively. Let the rank of AB equal the rank of A. Thus if $b_j = 0$, the corresponding $a_j = 0$. Further let Q_1 and Q_2 be linearly independent.

COROLLARY. *If the above conditions are satisfied, a necessary and sufficient condition that Q_2 and Q_1/Q_2 be stochastically independent is that $B^2 = bB$, where b is a real nonzero constant.*

This corollary is essentially the theorem suggested by von Neumann's original argument.

4. Ratios of linear forms.

THEOREM 3. *Let x have a continuous distribution such that $m(t) = E[\exp(tx)]$ exists for $-T < t < T$, $T > 0$. Let the real linear forms $L_1 = \sum_1^n a_j x_j$ and $L_2 = \sum_1^n x_j$, in n random values of x, be linearly independent. Provided $P(x \leqq 0) = 0$ $[P(x \geqq 0) = 0]$, a necessary and sufficient condition for L_2 and L_1/L_2 to be stochastically independent is that x $[-x]$ have a gamma distribution.*

PROOF OF SUFFICIENCY. We use Theorem 2. If x has a gamma distribution and the set $x_1, x_2, \cdots, x_n$ is a random sample, then $d_1 = d_2 = \cdots = d_n$. We also note that $b_1 = b_2 = \cdots = b_n = 1$. Hence $b_1 d_1 = b_2 d_2 = \cdots = b_n d_n$. This implies that L_2 and L_1/L_2 are stochastically independent.

PROOF OF NECESSITY. Write

$$M(u, t) = E[\exp(uL_1 + tL_2)],$$
$$= \prod_1^n m(a_j u + t).$$

Since the conditions of Theorem 1 are satisfied, the stochastic independence of L_2 and L_1/L_2 implies

$$\frac{\partial^k M(0, t)}{\partial u^k} \equiv K_k \frac{\partial^k M(0, t)}{\partial t^k}, \qquad k = 0, 1, 2, \cdots. \tag{4.1}$$

Using this condition for $k = 1$ we find

$$\sum_1^n a_j = nK_1. \tag{4.2}$$

For $k = 2$, (4.1) becomes

$$\begin{aligned}\left(\sum_1^n a_j^2\right)[m''(t)][m(t)]^{n-1} &+ \left(2\sum_{i<j} a_i a_j\right)[m'(t)]^2[m(t)]^{n-2} \\ &\equiv K_2\{n[m''(t)][m(t)]^{n-1} + n(n-1)[m'(t)]^2[m(t)]^{n-2}\}.\end{aligned} \tag{4.3}$$

We now show that this identity implies that

$$[m''(t)][m(t)]^{n-1} = c[m'(t)]^2[m(t)]^{n-2}, \tag{4.4}$$

where

$$c = \frac{m''(0)[m(0)]^{n-1}}{[m'(0)]^2[m(0)]^{n-2}}.$$

To do this we assume (4.4) is not true. That is, we assume $m''(t)[m(t)]^{n-1}$ and $[m'(t)]^2[m(t)]^{n-2}$ to be linearly independent. By considering the coefficients of the linearly independent functions in (4.3), we find

$$\sum_1^n a_j^2 = nK_2$$

and

$$2 \sum_{i<j} a_i a_j = n(n-1)K_2.$$

Adding these two equations we have

$$\left(\sum_1^n a_j\right)^2 = n^2 K_2.$$

This result with (4.2) implies that $K_1^2 = K_2$. However $K_1 = E[L_1/L_2]$ and $K_2 = E[(L_1/L_2)^2]$; so the variance of the ratio must equal zero. This requires the ratio to equal a constant; that is, $K_1 = L_1/L_2$. However this is contrary to the hypothesis that L_1 and L_2 be linearly independent. Thus (4.4) must be an identity.

We have now found that the stochastic independence of L_2 and L_1/L_2 imposes the restriction

$$m''(t)\, m(t) = c[m'(t)]^2$$

on the moment generating function of the distribution from which the samples are drawn. Since $m(t)$ is a moment generating function, $m(0) = 1$, $m'(0) = E(x)$, and $m''(0) = E(x^2)$. Moreover, with a continuous distribution, $E(x^2) > [E(x)]^2$ and hence $c > 1$. Accordingly, we can say that (4.1) for $k = 1, 2$ requires $m(t)$ to be the unique solution to the above differential equation with the given boundary condition $m(0) = 1$. That is,

$$m(t) = (1 - bt)^{1/(1-c)}, \qquad c > 1,$$

where b is an arbitrary constant. Hence (4.1) for $k = 1, 2$ restricts us to moment generating functions of the gamma type. It might be urged that (4.1) for $k = 3, 4, 5 \cdots$ could further restrict our solution. But this can not be the case since we proved the sufficiency of the gamma distribution for the stochastic independence of L_2 and L_1/L_2. That is, $M(u,t)$ must satisfy (4.1) if $m(t) = E[\exp(tx)]$, where x has a gamma distribution. This completes the proof of the necessity of the condition.

The author wishes to express his appreciation to Professor A. T. Craig for the suggestions made during the preparation of this paper.

REFERENCES

[1] H. Weyl, *The Theory of Groups and Quantum Mechanics*, Methuen and Co., Ltd., London, 1931.

[2] Tjalling Koopmans, "Serial correlation and quadratic forms in normal variables," *Annals of Math. Stat.*, Vol. 13 (1942), pp. 14–33.

[3] J. von Neumann, "Distribution of the ratio of the mean square successive difference to the variance," *Annals of Math. Stat.*, Vol. 12 (1941), pp. 367–395.

[4] J. D. Williams, "Moments of the ratio of the mean square successive difference to the mean square difference in samples from a normal universe," *Annals of Math. Stat.*, Vol. 12 (1941), pp. 239–241.

[5] Robert V. Hogg, "On ratios of certain algebraic forms in statistics," unpublished thesis, State University of Iowa.

NORMAL REGRESSION THEORY IN THE PRESENCE OF INTRA-CLASS CORRELATION

By Max Halperin[1]

USAF School of Aviation Medicine[2]

1. Summary. In this paper we prove that certain estimators and tests of significance used in regression analysis when observations are independent are equally valid in the presence of intra-class correlation. An application of this result is presented for the situation in which several replications of the correlated set of observations are available. As a special case of this application, it is shown that the usual test of "column effects" in the analysis of variance for a two-way classification remains valid when rows are independent and columns are uniformly correlated. This latter fact is also pointed out in [3].

2. Introduction. In the usual treatment of regression theory, as in [1] (Chapters VIII and IX), it is assumed that we have a sample of n independent observations, $y_1, \cdots, y_n$, where y_α arises from a normal distribution with mean $\sum_{p=1}^{k} C_p x_{p\alpha}$, and variance σ^2. Here, the $x_{p\alpha}$ are taken to be fixed variates. On the basis of these assumptions, unbiased estimates of $C_1, C_2, \cdots, C_k$ are obtained, and two theorems are proved, one concerning the joint distribution of the estimates of the C_p and the sum of squares of deviations from regression, the other concerning tests of significance of the C_p.

Now, on the one hand, it may happen that the results given in [1] are applied when, unknown to the experimenter, the observations are actually correlated. On the other hand, it may be clear, a priori, that the observations are correlated and that estimates and tests of the C_p are required in the light of the particular kind of correlation assumed to hold. In either case an investigation of estimates and distributions is called for. We consider these questions in Section 3 for the case that $y_1, \cdots, y_n$ have a variance matrix

$$R_n = \sigma^2 \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \cdot & \cdot & & \\ \cdot & \cdot & & \\ \cdot & \cdot & & \\ \rho & \rho & \cdots & 1 \end{pmatrix}. \tag{2.1}$$

In Section 4 we consider an application of our result to several replications of the correlated set of observations.

3. Estimates and significance tests in normal regression theory for correlated observations. We slightly modify the regression model indicated in Section 2

[1] Now at National Heart Institute, Bethesda, Md.

[2] This paper represents the views of the author and not necessarily those of the Department of the Air Force.

by supposing that the expected values of the y_α are given by

$$Ey_\alpha = \mu + \sum_{p=1}^{k} C_p x_{p\alpha}, \qquad \alpha = 1, 2, \cdots, n. \tag{3.1}$$

The reason for this modification will be apparent later. Assuming then that the y_α have the covariance matrix, R_n, the appropriate sample likelihood of $y_1, \cdots, y_n$ is readily seen to be

$$p(y_1, \cdots, y_n) = \frac{| R_n |^{-1/2}}{(2\pi)^{n/2}} \exp - \frac{1}{2} \{y - Ey\} R_n^{-1} \{y - Ey\}', \tag{3.2}$$

where

$$\begin{aligned} y &= (y_1, \cdots, y_n), \\ Ey &= \mu(1, \cdots, 1) + \sum_{p=1}^{k} C_p(x_{p1}, \cdots, x_{pn}). \end{aligned} \tag{3.21}$$

The maximum likelihood equations for the estimation of parameters from (3.2) are of such a formidable character that an explicit solution does not appear possible. As alternative estimates for μ, $C_1, \cdots, C_k$, one can use

$$\begin{aligned} \hat{\mu} &= \bar{y} - \sum_{p=1}^{k} \hat{C}_p \bar{x}_p, \\ \hat{C}_p &= \sum_{r=1}^{k} s_{ry} S_{rp}, \qquad p = 1, 2, \cdots, k, \end{aligned} \tag{3.3}$$

where

$$s_{ry} = \sum_{\alpha=1}^{n} (x_{r\alpha} - \bar{x}_r)(y_\alpha - \bar{y}), \qquad r = 1, 2, \cdots, k, \tag{3.31}$$

and the S_{rp} are elements of the inverse of

$$S = \begin{Vmatrix} s_{11} & \cdots & s_{1k} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ s_{k1} & \cdots & s_{kk} \end{Vmatrix}, \tag{3.32}$$

where

$$s_{ij} = \sum_{\alpha=1}^{n} (x_{i\alpha} - \bar{x}_i)(x_{j\alpha} - \bar{x}_j), \qquad i, j = 1, 2, \cdots, k. \tag{3.33}$$

We go on now to investigate the distribution of $\hat{C}_1, \cdots, \hat{C}_p$, when (3.2) holds. We have the following

THEOREM A. *Let $y_1, \cdots, y_n$ be a sample of one from a multivariate normal population with covariance matrix R_n and means $\mu + \sum_{p=1}^{k} C_p x_{p\alpha}$, $\alpha = 1, \cdots, n$. Let estimates of μ and the C_p be $\hat{\mu}$ and the $\hat{C}_p$ as defined in* (3.3). *Then*

(*a*) *the* $(\hat{C}_p - C_p)$ *have a multivariate normal distribution with zero means and covariance matrix* $(1-\rho)\sigma^2 S^{-1}$, *and*

(*b*) *the quantity* $\sum_{\alpha=1}^{n}(y_\alpha - \hat{\mu} - \sum_{p=1}^{k}\hat{C}_p x_{p\alpha})^2 (=V)$ *is distributed as* $(1-\rho)\sigma^2\chi^2$ *with* $(n-k-1)$ *degrees of freedom, and independently of the* $\hat{C}_p$.

PROOF. Conclusion (a) of the theorem follows readily from the fact that the $\hat{C}_p$ are linear functions of variables obeying a multivariate normal law and from some simple calculations to verify that the $\hat{C}_p$ are unbiased and have the indicated covariance matrix. The details are omitted.

Now let

$$L_n = \begin{pmatrix} l_{11} & \cdots & l_{1,n-1} & \frac{1}{\sqrt{n}} \\ \cdot & & \cdot & \cdot \\ \cdot & & \cdot & \cdot \\ \cdot & & \cdot & \cdot \\ l_{n1} & \cdots & l_{n,n-1} & \frac{1}{\sqrt{n}} \end{pmatrix}$$

be an $n \times n$ orthogonal matrix, and let

$$\begin{aligned} z &= yL_n, \\ w_p &= x_p L_n, \qquad p = 1, 2, \cdots, k \end{aligned} \tag{3.5}$$

By this transformation (3.2) becomes

$$\begin{aligned} p(z_1, \cdots, z_n) = {} & \frac{1}{\{2\pi\sigma^2(1-\rho)\}^{n-1/2}} \exp\left[-\frac{1}{2\sigma^2(1-\rho)}\sum_{\alpha=1}^{n-1}(z_\alpha - Ez_\alpha)^2\right] \\ & \cdot \frac{1}{\sigma\sqrt{2\pi}\,\{1+(n-1)\rho\}^{1/2}} \exp\left[-\frac{(z_n - Ez_n)^2}{2\sigma^2\{1+(n-1)\rho\}}\right] \end{aligned} \tag{3.51}$$

while

$$\begin{aligned} s_{ij} &= \sum_{\alpha=1}^{n} w_{i\alpha} w_{j\alpha} - w_{in} w_{jn} \\ &= \sum_{\alpha=1}^{n-1} w_{i\alpha} w_{j\alpha}, \\ s_{ry} &= \sum_{\alpha=1}^{n-1} w_{r\alpha} z_\alpha = s_{rz}. \end{aligned}$$

Applying the transformation (3.5) to the $\hat{C}_p$ and V, it is easy to show that

$$\begin{aligned} \hat{C}_p &= \sum_{r=1}^{k} s_{rz} S_{rp}, \\ V &= \sum_{\alpha=1}^{n-1}\left(z_\alpha - \sum_{p=1}^{k}\hat{C}_p w_{p\alpha}\right)^2. \end{aligned}$$

Since it can also be shown that

$$Ez_\alpha = \sum_{p=1}^{k} C_p w_{p\alpha}, \qquad \alpha = 1, 2, \cdots, n-1,$$

it is clear that the transformation (3.5) has reduced the problem to the standard one indicated in Section 2, with $(n-1)$ variables instead of n, and the theorem follows by the arguments given in [1].

THEOREM B. *Let $y_1, \cdots, y_n$ be as specified in Theorem A. Let H_0 be the statistical hypothesis that $C_{r+1} = C_{r+1,0}, \cdots, C_k = C_{k,0}$, regardless of the values of $C_1, \cdots, C_r$. When H_0 is true, the quantities*

$$V = \sum_{\alpha=1}^{n} \left(y_\alpha - \hat{\mu} - \sum_{p=1}^{k} \hat{C}_p x_{p\alpha} \right)$$

and

$$q = \sum_{g,h=r+1}^{k} b_{gh}(\hat{C}_g - C_{g,0})(\hat{C}_h - C_{h,0})$$

are independently distributed as $(1-\rho)\sigma^2\chi^2$ with $(n-k-1)$ and $(k-r)$ degrees of freedom respectively. Here $\hat{C}_p$ is defined by (3.3) and the b_{gh} are defined by the matrix equation

$$\begin{pmatrix} b_{r+1,r+1} & \cdots & b_{r+1,k} \\ \cdot & & \\ \cdot & & \\ \cdot & & \\ b_{k,r+1} & \cdots & b_{kk} \end{pmatrix} = \begin{pmatrix} S_{r+1,r+1} & \cdots & S_{r+1,k} \\ \cdot & & \\ \cdot & & \\ \cdot & & \\ S_{k,r+1} & \cdots & S_{kk} \end{pmatrix}^{-1}. \tag{3.6}$$

Also

$$F = \frac{(n-k-1)\, q}{(k-r)\, V}$$

provides a test of H_0 for $1 > \rho > -1/n-1$.

PROOF. It is clear that application of the transformation (3.5) will reduce the problem to that of proving the corresponding theorem in standard regression theory with a sample of $(n-1)$ independent observations. The theorem follows.

4. An application. We suppose we have m replications of the correlated sample of Section 3, generalizing slightly by further assuming that μ differs from replication to replication, assuming the value r_i for the ith replication. Thus, if $y_{i\alpha}$ is the αth measurement in the ith replication,

$$Ey_{i\alpha} = r_i + \sum_{p=1}^{k} C_p x_{p\alpha}, \qquad i = 1, 2, \cdots, m, \quad \alpha = 1, 2, \cdots, n, \tag{4.1}$$

and we ask for estimates of the r_i and C_p, and tests of significance for the C_p.

It follows easily from Section 3 that unbiased estimates of the r_i and the C_p are given by

$$(4.2)\qquad \begin{aligned} r_i &= \bar{y}_{i\cdot} - \sum_{p=1}^{k} \hat{C}_p \bar{x}_p, & i &= 1, 2, \cdots, m, \\ \hat{C}_p &= \sum_{r=1}^{k} s_{r\bar{y}} S_{rp}, & p &= 1, 2, \cdots, k, \end{aligned}$$

where

$$\begin{aligned} s_{r\bar{y}} &= \sum_{\alpha=1}^{n} (x_{r\alpha} - \bar{x}_r)(\bar{y}_{\cdot\alpha} - \bar{y}_{\cdot\cdot}), \\ \bar{y}_{i\cdot} &= \frac{1}{n} \sum_{\alpha=1}^{n} y_{i\alpha}, \\ \bar{y}_{\cdot\alpha} &= \frac{1}{m} \sum_{i=1}^{m} y_{i\alpha}, \\ \bar{y}_{\cdot\cdot} &= \frac{1}{mn} \sum_{i=1}^{m} \sum_{\alpha=1}^{n} y_{i\alpha}, \end{aligned}$$

and S_{rp} is defined as in (3.33).

We now ask for the joint distribution of $\hat{C}_1, \cdots, \hat{C}_k$ and

$$(4.22)\qquad V = \sum_{i=1}^{m} \sum_{\alpha=1}^{n} \left\{ y_{i\alpha} - \bar{y}_{i\cdot} - \sum_{p=1}^{k} \hat{C}_p (x_{p\alpha} - \bar{x}_p) \right\}^2.$$

It follows as in Section 3 that $\hat{C}_1, \cdots, \hat{C}_k$, have a multivariate normal distribution, and it is sufficient for our purposes to examine the joint distribution of V and W, where

$$W = m(\hat{C} - C)S(\hat{C} - C)', \qquad \hat{C} - C = (\hat{C}_1 - C_1, \cdots, \hat{C}_k - C_k).$$

By an application of the transformation $z_i = y_i L_n$ to the n observations of each replication, one obtains

THEOREM A′. *Let $y_{i1}, \cdots, y_{in} (i = 1, 2, \cdots, m)$ be a sample of one from a multivariate normal population with means given by* (4.1) *and the $mn \times mn$ variance matrix*

$$R_{nm} = \begin{vmatrix} R_n & 0 & \cdots & 0 \\ 0 & R_n & \cdots & 0 \\ \cdot & \cdot & & \\ \cdot & \cdot & & \\ \cdot & \cdot & & \\ 0 & 0 & \cdots & R_n \end{vmatrix}.$$

Then $(\hat{C}_1 - C_1), \cdots, (\hat{C}_k - C_k)$, have a multivariate normal distribution with zero means and variance matrix $[(1 - \rho)\sigma^2/m]S^{-1}$, and W and V are independent $(1 - \rho)\sigma^2\chi^2$ variates with k and $m(n - 1) - k$ degrees of freedom respectively.

We can also prove

THEOREM B′. *Let* $y_{i1}, \cdots, y_{in}(i = 1, 2, \cdots, m)$ *satisfy the conditions of Theorem A′. Let* H_0 *be as specified in Theorem B. Then the statistic*

$$F = \frac{[m(n-1) - k]\, q}{(k-r)\, V},$$

where

$$q = \sum_{g,h=r+1}^{k} b_{gh}(\hat{C}_g - C_{g,0})(\hat{C}_h - C_{h,0})$$

and b_{gh} *is defined by the matrix equation*

$$\begin{pmatrix} b_{r+1,r+1} & \cdots & b_{r+1,k} \\ \cdot & & \\ \cdot & & \\ \cdot & & \\ b_{k,r+1} & \cdots & b_{kk} \end{pmatrix} = m \begin{pmatrix} S_{r+1,r+1} & \cdots & S_{r+1,k} \\ \cdot & & \\ \cdot & & \\ \cdot & & \\ S_{k,r+1} & \cdots & S_{kk} \end{pmatrix}^{-1},$$

is distributed as Snedecor's F and provides a test of H_0.

The proof of Theorem B′ is along the same lines as that of Theorem A′ and is omitted.

We also remark that theorems akin to A′ and B′ hold if $r_i = r, i = 1, 2, \cdots, m$. We simply may take

$$\hat{r} = \bar{y}_{..} - \sum_{p=1}^{k} \hat{C}_p \bar{x}_p .$$

The estimates of the C_p need not be changed. If now we let

$$V' = \sum_{i=1}^{m} \sum_{j=1}^{n} \left\{ y_{ij} - \bar{y}_{..} - \sum_{p=1}^{k} \hat{C}_p (x_{pj} - \bar{x}_p) \right\}^2,$$

Theorems A′ and B′ hold with r_i and $\hat{r}_i$ replaced by r and $\hat{r}$, with V replaced by V' and $m(n-1) - k$ degrees of freedom replaced by $nm - k - 1$ degrees of freedom.

As an example of the application of these notions we consider an analysis of variance problem. The same problem has been considered in [3]. Suppose we have mn observations,

$$\begin{matrix} y_{11} & \cdots & y_{1n} \\ \cdot & & \\ \cdot & & \\ \cdot & & \\ y_{m1} & \cdots & y_{mn} \end{matrix}$$

where the $y_{i\alpha}$ are jointly normal with covariance matrix R_{nm} and with means given by

$$Ey_{i\alpha} = r_i + C_\alpha . \tag{4.3}$$

In [3] it is shown that the F ratio for "columns" calculated in the usual way has the usual F distribution when the C_j are equal. To deduce this test from our results we write (4.3) as

$$Ey_{i\alpha} = r_i + \sum_{p=1}^{n} C_p x_{p\alpha}, \tag{4.31}$$

where

$$\begin{aligned} x_{p\alpha} &= 0, \qquad p \neq \alpha, \\ &= 1, \qquad p = \alpha. \end{aligned}$$

We have then

$$\begin{aligned} s_{pq} &= \sum_{\alpha=1}^{n} (x_{p\alpha} - \bar{x}_p)(x_{q\alpha} - \bar{x}_q) \\ &= -\frac{1}{n},\ p \neq q, \\ &= \frac{n-1}{n},\ p = q. \end{aligned}$$

The $n \times n$ matrix, S, is singular. To overcome this difficulty we can, since we are only interested in class differences rather than in the absolute values of the C_p, arbitrarily assign to one of the C_p, say C_n, the value zero. The test of column differences then becomes a test that $C_1 = C_2 = \cdots = C_{n-1} = 0$. It is then easy to see that $\hat{C}_p = \bar{y}_{\cdot p} - \bar{y}_{\cdot n}$, $p = 1, 2, \cdots, n-1$, and

$$\hat{r}_i = \bar{y}_{i\cdot} - \frac{1}{n} \sum_{p=1}^{n-1} \hat{C}_p .$$

If we substitute these values in $q = m\, \hat{C}\, S^*\, \hat{C}'$ and

$$V = \sum_{i=1}^{m} \sum_{\alpha=1}^{n} (y_{i\alpha} - \hat{r}_i - \sum_{p=1}^{n-1} \hat{C}_p x_{p\alpha})^2,$$

where

$$\hat{C} = (\hat{C}_1, \cdots, \hat{C}_{n-1})$$

and S^* is the minor of s_{nn} in S, we find after a little algebraic reduction that

$$F = \frac{(n-1)(m-1)q}{(n-1)V} = \frac{m(n-1)(m-1) \sum_{j=1}^{n} (\bar{y}_{\cdot j} - \bar{y}_{\cdot\cdot})^2}{(n-1) \sum_{i=1}^{m} \sum_{j=1}^{n} (y_{ij} - \bar{y}_{i\cdot} - \bar{y}_{\cdot j} + \bar{y}_{\cdot\cdot})^2},$$

and this is the desired statistic.

Suggestions of the referee for simplifying the proofs are gratefully acknowledged.

REFERENCES

[1] S. S. WILKS, *Mathematical Statistics*, Princeton University Press, 1946.
[2] H. CRAMÉR, *Mathematical Methods of Statistics*, Princeton University Press, 1946, pp. 490–496.
[3] D. F. VOTAW, A. W. KIMBALL, AND J. A. RAFFERTY, "Compound symmetry tests in the multivariate analysis of medical experiments," *Biometrics*, Vol. 6 (1950), pp. 259–281.
[4] J. E. WALSH, "Concerning the effect of intra-class correlation on certain significance tests," *Annals of Math. Stat.*, Vol. 18 (1947), pp. 88–96.

MINIMUM VARIANCE ESTIMATION WITHOUT REGULARITY ASSUMPTIONS

By Douglas G. Chapman[1] and Herbert Robbins

University of Washington and University of North Carolina

1. Summary and Introduction. Following the essential steps of the proof of the Cramér-Rao inequality [1, 2] but avoiding the need to transform coordinates or to differentiate under integral signs, a lower bound for the variance of estimators is obtained which is (a) free from regularity assumptions and (b) at least equal to and in some cases greater than that given by the Cramér-Rao inequality. The inequality of this paper might also be obtained from Barankin's general result[2] [3]. Only the simplest case—that of unbiased estimation of a single real parameter—is considered here but the same idea can be applied to more general problems of estimation.

2. Lower bound. Let μ be a fixed measure on Euclidean n-space X and let the random vector $x = (x_1, \cdots, x_n)$ have a probability distribution which is absolutely continuous with respect to μ, with density function $f(x, \alpha)$, where α is a real parameter belonging to some parameter set A. Define $S(\alpha)$ as follows:

$$f(x, \alpha) > 0, \qquad \text{a.e. } x \text{ in } S(\alpha),$$
$$f(x, \alpha) = 0, \qquad \text{a.e. } x \text{ in } X - S(\alpha).$$

Let $t = t(x)$ be any unbiased estimator of α, so that for every α in A,

$$\int_X tf(x, \alpha)\, d\mu = \alpha. \tag{1}$$

If α, $\alpha + h (h \neq 0)$ are any two distinct values in A such that

$$S(\alpha + h) \subset S(\alpha), \tag{2}$$

then, writing S for $S(\alpha)$,

$$\int_S f(x, \alpha)\, d\mu = 1, \qquad \int_{S(\alpha+h)} f(x, \alpha + h)\, d\mu = \int_S f(x, \alpha + h)\, d\mu = 1,$$

$$\int_S tf(x, \alpha)\, d\mu = \alpha, \qquad \int_S tf(x, \alpha + h)\, d\mu = \alpha + h,$$

so that

$$\int_S [t - \alpha] \sqrt{f(x, \alpha)}\, \frac{f(x, \alpha + h) - f(x, \alpha)}{hf(x, \alpha)} \sqrt{f(x, \alpha)}\, d\mu = 1.$$

---

[1] This research was supported in part by the Office of Naval Research.

[2] But again with some additional restrictions.

Applying Schwarz's inequality we obtain the relation

$$\begin{aligned}1 &\le \int_S [t-\alpha]^2 f(x,\alpha)\,d\mu \cdot \int_S \left[\frac{f(x,\alpha+h)-f(x,\alpha)}{hf(x,\alpha)}\right]^2 f(x,\alpha)\,d\mu \\ &= \operatorname{Var}(t \mid \alpha)\cdot \frac{1}{h^2}\left\{\int_S \left[\frac{f(x,\alpha+h)}{f(x,\alpha)}\right]^2 f(x,\alpha)\,d\mu - 1\right\}.\end{aligned} \tag{3}$$

Let

$$J = J(\alpha, h) = \frac{1}{h^2}\left\{\left[\frac{f(x,\alpha+h)}{f(x,\alpha)}\right]^2 - 1\right\};$$

then (3) can be written in the form

$$\operatorname{Var}(t \mid \alpha) \ge \frac{1}{E(J \mid \alpha)}. \tag{4}$$

Since (4) holds whenever α, $\alpha + h$ are any two distinct elements of A satisfying (2) we obtain the fundamental inequality

$$\operatorname{Var}(t \mid \alpha) \ge \frac{1}{\inf\limits_h E(J \mid \alpha)}, \tag{5}$$

where the infimum is taken over all $h \neq 0$ such that (2) is satisfied. It should be noted that (5) holds without any restriction on $f(x, \alpha)$ and without any restriction on t other than (1).

It is possible that $E(J \mid \alpha)$ does not exist (finitely) for any h. With the usual convention that $E(J \mid \alpha) = \infty$, in this case, (5) is still a valid, though trivial, inequality.

In applications μ will often be Lebesgue measure on X. It could equally well be a discrete measure on a countable set of points in X. Furthermore, if the set where $f(x, \alpha) > 0$ is independent of α then (2) is trivially satisfied for all $\alpha + h$ in A.

We shall have occasion to compare (5) with the Cramér-Rao inequality

$$\operatorname{Var}(t \mid \alpha) \ge \frac{1}{E(\psi^2 \mid \alpha)}; \qquad \psi = \psi(\alpha) = \frac{\partial}{\partial \alpha} \ln f(x, \alpha). \tag{6}$$

This inequality is usually derived for distributions with range independent of the parameter and under certain regularity conditions on both $f(x, \alpha)$ and the unbiased estimator t.

3. Examples.

Example 1. *Unbiased estimation of the mean of a normal distribution based on a random sample of size n.* Here

$$f(x, \alpha) = (2\pi)^{-(n/2)}\sigma^{-n}e^{-(1/2\sigma^2)\sum_{i=1}^n (x_i-\alpha)^2},$$

where σ is a positive constant, and

$$J = \frac{1}{h^2}\{e^{-(1/\sigma^2)\sum_{i=1}^n [(x_i-\alpha-h)^2-(x_i-\alpha)^2]} - 1\} = \frac{n}{\sigma^2 k^2}\{e^{-k^2}e^{2ku} - 1\},$$

where we have set $u = \sum_{i=1}^{n}(x_i - \alpha)/(\sigma\sqrt{n})$, $\quad k = h\sqrt{n}/\sigma \neq 0$.

When the mean is α, u is normally distributed with mean 0 and variance 1, and we find after a simple computation that

$$(7)\qquad \begin{aligned} E(J \mid \alpha) &= n(e^{k^2} - 1)/(\sigma^2 k^2), \\ \inf_h E(J \mid \alpha) &= \lim_{k \to 0} [n(e^{k^2} - 1)/(\sigma^2 k^2)] = n/\sigma^2 = [E(\psi^2 \mid \alpha)]. \end{aligned}$$

Hence if t is any unbiased estimator of α it follows from (5) that

$$(8)\qquad \operatorname{Var}(t \mid \alpha) \geq \sigma^2/n.$$

Since the sample mean $\bar{x}$ is an unbiased estimator of α with $\operatorname{Var}(\bar{x} \mid \alpha) = \sigma^2/n$, it follows that $\bar{x}$ has minimum variance in the class of *all* unbiased estimators of α.

In this example the Cramér-Rao inequality (6) yields precisely the same bound (8).

Corresponding results hold for the unbiased estimation of the variance when the mean is known. Both (5) and (6) yield the inequality

$$\operatorname{Var}(t \mid \alpha) \geq 2\alpha^2/n,$$

where α is the unknown variance. The equality sign holds for

$$t = n^{-1} \sum_{i=1}^{n} (x_i - m)^2,$$

where m is the mean of the normal population.

Example 2. Unbiased estimation of the standard deviation of a normal population with known mean. Here

$$f(x, \alpha) = (2\pi)^{-(n/2)} \alpha^{-n} e^{-(1/2\alpha^2)\sum_{i=1}^{n}(x_i - m)^2}.$$

Setting $k = h/\alpha$ we find that for $-1 < k < \sqrt{2} - 1$, $k \neq 0$,

$$(9)\qquad E(J \mid \alpha) = \{(1 + k)^{-n}[1 - k(2 + k)]^{-(n/2)} - 1\}/(\alpha^2 k^2).$$

In this case also, $\lim_{k\to 0} E(J \mid \alpha) = 2n/\alpha^2 = E(\psi^2 \mid \alpha)$. But the minimum value of $E(J \mid \alpha)$ is not approached in the neighborhood of $h = k = 0$, and the inequality (5) is sharper than (6). We shall consider only the case $n = 2$. Equation (9) then becomes

$$E(J \mid \alpha) = (p + 1)^2/[\alpha^2 p^2 (2 - p^2)],$$

where we have set $p = 1 + k$ and $0 < p < \sqrt{2}$. We have for $p = .8393$, $1/E(J \mid \alpha) = .2698\ \alpha^2$, so that by (5)

$$\operatorname{Var}(t \mid \alpha) \geq .2698\ \alpha^2 > .25\alpha^2 = \frac{1}{E(\psi^2 \mid \alpha)}.$$

It is interesting to note that the unbiased estimator

$$t = \sqrt{\tfrac{1}{2}n}\,\frac{\Gamma(\frac{1}{2}n)}{\Gamma[\frac{1}{2}(n+1)]}\sqrt{\sum_{i=1}^{n}(x_i - m)^2 \Big/ n}$$

has variance

$$\alpha^2\left[\tfrac{1}{2}n\,\frac{\Gamma^2\frac{1}{2}n}{\Gamma^2[\frac{1}{2}(n+1)]} - 1\right],$$

which for $n = 2$ becomes

$$\alpha^2\left[\frac{4}{\pi} - 1\right] = .2732\alpha^2.$$

But it can be shown using results of Lehmann and Scheffé [4], or of Hoel [5], which were derived from Blackwell's theorem on conditional expectation [6], that no other unbiased estimator can have smaller variance than t. Thus (5) does not give the *greatest* lower bound in this case.

Various examples of the application of (5) can be given where $S(\alpha)$ is not a constant and where the Cramér-Rao formula is invalid (see for example Cramér [1], p. 485). It should be noted, however, that in many of the standard problems of this type stronger results can be obtained by other methods.

Another class of estimation problems where (5) may be applied occurs if the parameter space is discrete. Again in this case the Cramér-Rao formula does not hold. An example of this type has been given by Chapman ([7], pp. 149–150). Other applications of this type and some results related to this paper were obtained recently by Hammersley [8].

4. General comparison with the Cramér-Rao inequality. Let

$$\mathfrak{J} = \mathfrak{J}(\alpha, h) = \left[\frac{f(x, \alpha + h) - f(x, \alpha)}{hf(x, \alpha)}\right]^2; \tag{10}$$

then

$$E(\mathfrak{J} \mid \alpha) = E(J \mid \alpha).$$

Hence in the fundamental inequality (5) we can replace J by $\mathfrak{J}$. But from (10) it is clear that

$$\lim_{h\to 0} \mathfrak{J}(\alpha, h) = \left[\frac{\partial}{\partial\alpha}\ln f(x, \alpha)\right]^2 = \psi^2(\alpha)$$

whenever the latter exists.

Assuming now the usual regularity conditions under which the Cramér-Rao lower bound is derived, that $S(\alpha)$ is independent of α and that $f(x, \alpha)$ is sufficiently regular that we may pass to the limit inside the integral sign,

$$E(\psi^2 \mid \alpha) = E\left[\lim_{h\to 0}(\mathfrak{J} \mid \alpha)\right] = \lim_{h\to 0} E(\mathfrak{J} \mid \alpha) \geq \inf_h E(\mathfrak{J} \mid \alpha) = \inf_h E(J \mid \alpha), \tag{11}$$

the infimum being taken over admissible values of h. It follows that *the inequality* (5) *is at least as sharp as that given by the Cramér-Rao formula* (6).

On the other hand, when $x = (x_1, \cdots, x_n)$ is a random sample from a regular distribution, and when $E(\psi^2 \mid \alpha) < \infty$, then for any *fixed* $h \neq 0$, there exists an n_0 such that for $n > n_0$

$$E(\psi^2 \mid \alpha) \leq E(J \mid \alpha). \tag{12}$$

Without loss of generality assume $E(J \mid \alpha) < \infty$. Letting $g(t, \alpha)$ denote the density function of a single x_i and ν the one-dimensional measure which generates μ, it is easily verified that

$$E(J \mid \alpha) = \frac{1}{h^2}\left(\left[\int_{\mathfrak{X}} \frac{g^2(t, \alpha + h)}{g(t, \alpha)}\, d\nu\right]^n - 1\right).$$

By hypothesis, except on a set of measure 0,

$$g(t, \alpha + h) = g(t, \alpha) + h \left.\frac{\partial g}{\partial \alpha}\right|_{\alpha=\alpha(h)}; \qquad \alpha \leq \alpha(h) \leq \alpha + h.$$

Hence

$$\int_{\mathfrak{X}} \frac{g^2(t, \alpha + h)}{g(t, \alpha)}\, d\nu = 1 + 2h \int_{\mathfrak{X}} \left.\frac{\partial g}{\partial \alpha}\right|_{\alpha=\alpha(h)} d\nu + h^2 \int_{\mathfrak{X}} g^{-1}\left(\left.\frac{\partial g}{\partial \alpha}\right|_{\alpha=\alpha(h)}\right)^2 d\nu. \tag{13}$$

Denoting the last integral of the right hand side of (13) by $R(\alpha, h)$ and noting that the relation

$$\int_{\mathfrak{X}} g(t, \alpha)\, d\nu = 1$$

may be differentiated under the integral sign so that the middle term vanishes, it follows that

$$E(J \mid \alpha) = \frac{[1 + h^2 R(\alpha, h)]^n - 1}{h^2} \geq nR(\alpha, h) + \tfrac{1}{2}n(n-1)h^2 R^2(\alpha, h). \tag{14}$$

On the other hand, from (11) and (14),

$$E(\psi^2 \mid \alpha) = nR(\alpha, 0). \tag{15}$$

In order that different parameters may be distinguishable we must have

$$\left.\frac{\partial g}{\partial \alpha}\right|_{\alpha=\alpha(h)} \neq 0$$

for a set of positive measure on the t-axis, and hence $R(\alpha, h) > 0$. From this and the fact that $R(\alpha, 0)$ is independent of n, (12) follows at once, for sufficiently large n, from (14) and (15).

REFERENCES

[1] H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, 1946.

[2] C. R. Rao, "Information and the accuracy attainable in the estimation of statistical parameters," *Bull. Calcutta Math. Soc.*, Vol. 37 (1945), pp. 81–91.

[3] E. W. BARANKIN, "Locally best unbiased estimates," *Annals of Math. Stat.*, Vol. 20 (1949), pp. 477–501. (More complete references to the general problem are given in this paper.)

[4] E. L. LEHMANN AND H. SCHEFFÉ, "Completeness, similar regions and unbiased estimation, Part 1," *Sankhyā*, Vol. 10 (1950), pp. 305–340.

[5] P. G. HOEL, "Conditional expectation and the efficiency of estimates," *Annals of Math. Stat.*, Vol. 22 (1951), pp. 299–301.

[6] D. BLACKWELL, "Conditional expectation and unbiased sequential estimation," *Annals of Math. Stat.*, Vol. 18 (1947), pp. 105–110.

[7] D. G. CHAPMAN, "Some properties of the hypergeometric distribution with applications to zoological sample censuses," *Univ. of California Publ. Statist.*, Vol. 1, No. 7 (1951), pp. 131–160.

[8] J. M. HAMMERSLEY, "On estimating restricted parameters," *Jour. Roy. Stat. Soc., Ser. B*, Vol. 12 (1950), pp. 192–229.

NOTES

A GENERAL CONCEPT OF UNBIASEDNESS

By E. L. Lehmann

University of California, Berkeley, and Princeton University

The term unbiasedness was introduced by Neyman and Pearson [1] in connection with hypothesis testing. A test of the hypothesis $\theta \varepsilon \omega$ against the alternatives $\theta \varepsilon \Omega - \omega$ is said to be unbiased at level α if its power function β satisfies

$$(1) \qquad \begin{aligned} \beta(\theta) &\leq \alpha \text{ for } \theta \,\varepsilon\, \omega, \\ \beta(\theta) &\geq \alpha \text{ for } \theta \,\varepsilon\, \Omega - \omega. \end{aligned}$$

In 1937 Neyman [2] developed a theory of estimation by confidence sets. He established a duality with the theory of hypothesis testing, so that to each notion of one theory corresponds an analogous one in the other. In particular, he defined a family of confidence sets $A(x)$ to be unbiased if

$$(2) \qquad P_\theta(A(X) \supset \theta') \leq P_\theta(A(X) \supset \theta) \text{ for all } \theta \text{ and } \theta'.$$

While the above two definitions are closely related, a third use of the term unbiasedness was made in a rather different context. In presenting their version of the Gauss-Markov theorem on least squares David and Neyman [3] defined a point estimate $\delta(X)$ of $g(\theta)$ to be unbiased if its expectation coincides with the estimated value, that is, if

$$(3) \qquad E_\theta\delta(X) \equiv g(\theta).$$

It was pointed out later by Brown [4] that one obtains other analogous definitions by postulating that some central value of the distribution of $\delta(X)$ other than the mean coincides with the estimated value. Using the median as an example he defined $\delta(X)$ to be median-unbiased if

$$(4) \qquad P_\theta(\delta(X) > g(\theta)) = P_\theta(\delta(X) < g(\theta)) \text{ for all } \theta.$$

In view of Wald's theory of decision functions [5] it seems tempting to try to give a definition of unbiasedness at the level of generality of this theory. Suppose we are concerned with a decision problem where the loss resulting from a decision $\delta(X)$ is $W(\theta, \delta(X))$ when the true parameter value is θ. In analogy with (2) we shall say that a decision procedure $\delta(X)$ is unbiased if for each θ

$$(5) \qquad E_\theta W(\theta', \delta(X)) = \min \text{ when } \theta' = \theta.$$

This clearly reduces to Neyman's definition for confidence sets if one uses for loss function,

$$(6) \qquad W(\theta, \delta(x)) = \begin{cases} 0 \text{ if the confidence set } \delta(x) \text{ covers } \theta, \\ 1 \text{ otherwise.} \end{cases}$$

In order to obtain an interpretation of condition (5), let us consider the case that for each parameter value θ there exists a unique "correct" decision d and that each d is correct for at least some θ. This is the case for example in hypothesis testing and in point estimation. Here a correct decision may be defined by the condition $W(\theta, d) = 0$. Let us say that two parameter values θ, θ' are equivalent, $\theta \sim \theta'$, if the correct decision is the same for both of them, and let us suppose that for any decision d'

$$W(\theta_1, d') = W(\theta_2, d') \text{ whenever } \theta_1 \sim \theta_2. \tag{7}$$

Then the loss $W(\theta, d')$ depends only on the actual decision taken, say d', and the decision d that would have been correct, and we may write for it $W(d, d')$. The loss $W(d, d')$ is a measure of how far the two decisions d and d' are apart, and (5) states that a decision function $\delta(X)$ is unbiased if on the average it comes closer to the correct decision than to any incorrect decision.

Let us now apply this notion to some particular examples. Let the decision to accept and reject the hypothesis $H: \theta \varepsilon \omega$ be denoted by d_0 and d_1, respectively. Since in the Neyman-Pearson theory of hypothesis testing one is concerned only with the probabilities of the two types of error, the natural associated loss function is of the form

$$\begin{aligned} W(\theta, d_0) &= \begin{cases} a \text{ if } \theta \in \Omega - \omega, \\ 0 \text{ if } \theta \in \omega; \end{cases} \\ W(\theta, d_1) &= \begin{cases} b \text{ if } \theta \in \omega, \\ 0 \text{ if } \theta \in \Omega - \omega. \end{cases} \end{aligned} \tag{8}$$

It is easy to see that in this case (5) becomes

$$\begin{aligned} P_\theta(d_1) &\leq \frac{a}{a+b} \text{ for } \theta \in \omega, \\ P_\theta(d_1) &\geq \frac{a}{a+b} \text{ for } \theta \in \Omega - \omega, \end{aligned} \tag{9}$$

where $P_\theta(d)$ denotes the probability that $\delta(X) = d$ when θ is the true parameter value. This is exactly the usual definition (1) with $\alpha = a/(a + b)$.

Let us next consider point estimation where the loss is taken as the square of the error. If the function to be estimated is $g(\theta)$, condition (5) becomes

$$E_\theta[\delta(X) - g(\theta')]^2 \geq E_\theta[\delta(X) - g(\theta)]^2 \text{ for all } \theta, \theta'. \tag{10}$$

Let $E_\theta\delta(X) = h(\theta)$. In the usual case that $h(\theta)$ is one of the possible values of the function g, the left-hand side of (10) is minimized for $g(\theta') = h(\theta)$. Thus the inequality holds for all θ' if and only if $g(\theta) = h(\theta)$, which is equivalent to (3). So again (5) reduces just to the usual definition.

Even if $h(\theta)$ is not one of the possible values of g, it is easily seen that (10) is equivalent to

$$| h(\theta) - g(\theta) | = \min_{\theta'} | h(\theta) - g(\theta') | .$$

Then, if for example Ω is a real interval and g is continuous and strictly monotone, there can exist at most two values of θ for which $g(\theta) \neq h(\theta)$. If further, as is usually the case, $h(\theta)$ is continuous for all estimates δ, we must have $h(\theta) \equiv g(\theta)$.

Quite analogously one sees that if $W(\theta, \delta(x)) = |\delta(x) - g(\theta)|$, definition (5) reduces to Brown's notion of median-unbiasedness.

While the definition given here seems satisfactory in that it does reduce under reasonable assumptions to the usual concepts, it is somewhat more restrictive than appears at first sight. If for example there exists for each θ a unique correct decision d and if the loss function is of the form

$$W(\theta, d') = f(\theta)V(d, d'),$$

then, with the trivial exception of procedures for which $E_\theta V(d, \delta(x)) = 0$ for some d and some value of θ in ω_d, no unbiased procedure can exist unless $f(\theta)$ is constant on each ω_d. For let $\theta, \theta' \in \omega_d$. On substituting in (5) we see that unbiasedness implies $f(\theta') \geq f(\theta)$ and hence by symmetry $f(\theta') = f(\theta)$. In hypothesis testing for example if the loss is zero for a correct decision, it follows, again with trivial exceptions, that unbiased tests can exist only if the loss function is given by (8).

It is perhaps worth pointing out certain connections between the principle of unbiasedness and that of invariance. Consider for example the problem of estimating θ from a sample $X_1, \cdots, X_n$ where the X's are uniformly distributed on $(0, \theta)$. If one takes as loss function

$$W(\theta, \delta(x_1, \cdots, x_n)) = [\delta(x_1, \cdots, x_n) - \theta]^2/\theta^2 \tag{11}$$

the problem transforms in an obvious manner under a change of scale, and one may wish to consider only estimates having the invariance property

$$\delta(cX_1, \cdots, cX_n) = c\delta(X_1, \cdots, X_n) \text{ for all } c > 0. \tag{12}$$

If $Y = \max(X_1, \cdots, X_n)$, it is easily seen that among all invariant estimates the one that uniformly minimizes the expected loss is

$$\frac{n+2}{n+1}Y. \tag{13}$$

This estimate does not have the usual unbiasedness property since

$$E_\theta\left[\frac{n+2}{n+1}Y\right] = \frac{n(n+2)}{(n+1)^2}\theta.$$

However a simple computation shows that (13) is unbiased in the sense of (5) with respect to the invariant loss function (11).

More generally, let $\mathcal{G}$ be a group of measurable 1:1 transformations on the sample space. Let gX be the random variable that takes on the value gx when $X = x$, and suppose that when X has a distribution p_θ, $\theta \in \Omega$, then gX has a distribution $p_{\theta'}$, $\theta' \in \Omega$. Denote this θ' by $\bar{g}\theta$ and suppose that $\bar{g}\theta$ defines a 1:1 transformation on Ω. Let $\bar{\mathcal{G}}$ be the group of transformations $\bar{g}$ and assume that

there exists a group $\mathcal{G}^*$ of 1:1 transformations on the decision space D such that $\mathcal{G}^*$ is homomorphic to $\bar{\mathcal{G}}$ and

$$W(\bar{g}\theta, g^*d) = W(\theta, d) \text{ for all } \theta \in \Omega, d \in D. \tag{14}$$

Then a decision function δ is said to be invariant if

$$\delta(gX) = g^*\delta(X). \tag{15}$$

This is a natural generalization of the definition of invariance given by Hunt and Stein [6, 7], and is essentially the definition used by Peisakoff [8]. Further, δ is said to be almost invariant if (15) holds except on a set N_g of measure 0.

Whenever among all unbiased procedures there exists a unique[1] one that uniformly minimizes the risk, then it is almost invariant. This follows easily from the fact that if $\delta(X)$ is unbiased $g^*\delta(g^{-1}X)$ is also unbiased. It is not in general true that conversely an optimum invariant test is necessarily unbiased. However, this result does hold under certain restrictions.[2] If

(i) $\mathcal{G}$ is transitive, i.e., given any θ, θ' there exists $\bar{g}$ such that $\theta = \bar{g}\theta'$,

(ii) $\mathcal{G}^*$ is commutative,

and if among all invariant (or almost invariant) procedures there exists one that uniformly minimizes the risk, then it is unbiased.

To see this, let δ be invariant and such that for any other invariant procedure δ'

$$E_\theta W(\theta, \delta'(X)) \geq E_\theta W(\theta, \delta(X)).$$

Let $\theta' \neq \theta$, $\theta = \bar{g}\theta'$, say. Then

$$E_\theta W(\theta', \delta(X)) = E_\theta W(\theta, g^*\delta(X)) \geq E_\theta W(\theta, \delta(X)).$$

Here the inequality follows since by (ii) the invariance of $\delta(X)$ implies that $g^*\delta(X)$ is also invariant.

While assumptions (i) and (ii) are satisfied in many estimation problems, (i) will in general not hold in a problem of hypothesis testing because of the asymmetry of d_0 and d_1. Here the result in question follows when the loss function is given by (8) from the fact that if a test is unbiased so is any test that is uniformly better, together with the unbiasedness and invariance of the test $\varphi(x) \equiv a/(a + b)$ (i.e., the test that rejects the hypothesis with probability $a/(a + b)$ regardless of the observations).

That the result is not true in general if we drop either one of the two conditions (i) or (ii) can be seen from the following example. For estimating the mean ξ of a normal variable with unknown variance σ^2 when the loss function is $[(\delta(x) - \xi)/\sigma]^2$, the best invariant estimate is X both with respect to the group

$$\mathcal{G}_1: gx = x + b, \qquad -\infty < b < \infty$$

[1] Throughout, this is understood to mean unique up to a set of measure zero.

[2] I am grateful to the referee for pointing out an error in my original statement of this result.

and with respect to

$$\mathcal{G}_2: gx = ax + b, \qquad 0 < a < \infty, -\infty < b < \infty.$$

For this problem an unbiased estimate in the sense of (5) does not exist, and it is seen that $\mathcal{G}_1$ satisfies (ii) but not (i) while $\mathcal{G}_2$ satisfies (i) but not (ii).

The notion of unbiasedness in many cases leads to reasonable decision procedures and this seems to be in general the value of such concepts. On the other hand there is no guarantee that an optimum unbiased procedure is necessarily satisfactory. As an example (for another example see [9]) consider a Poisson variable X which is observed only if $X \neq 0$, so that the distribution of X is given by

$$P(X = K) = \frac{\lambda^K}{K!} e^{-\lambda}(1 - e^{-\lambda})^{-1}, \qquad K = 1, 2, \cdots. \tag{15}$$

It is desired to estimate the probability $e^{-\lambda}$ of X being zero, and the loss function is squared error. The condition of unbiasedness gives

$$\sum_{K=1}^{\infty} \delta(K) \frac{\lambda^K}{K!} \equiv 1 - e^{-\lambda} \equiv \sum_{K=1}^{\infty} (-1)^{K+1} \frac{\lambda^K}{K!}, \tag{16}$$

so that $\delta(K) = (-1)^{K+1}$. Thus the estimate takes on only impossible values and instead of decreasing with K as one would expect, it does not depend on the order of magnitude of K at all.

As a final remark we mention, without going into details, the following extension of the notion of unbiasedness. Instead of comparing $E_\theta W(\theta', \delta(X))$ only with $E_\theta W(\theta, \delta(X))$ we may ask that $E_\theta W(\theta', \delta(X))$ be a nondecreasing function of $v(\theta, \theta')$, where $v(\theta, \theta')$ in some sense measures the distance between θ and θ'. This notion is a generalization of one used by P. L. Hsu [10] in the theory of hypothesis testing. It is also closely connected with the principle of invariance. In fact if there exists a group of transformations leaving the problem invariant then with a suitable definition of $v(\theta, \theta')$ it is easy to see under weak assumptions on the loss function that Theorem 7.1 of [7] generalizes to the present case. This theorem states essentially that the totality of procedures for which $E_\theta W(\theta', \delta(X))$ depends only on $v(\theta, \theta')$ coincides with the totality of invariant procedures.

REFERENCES

[1] J. Neyman and E. S. Pearson, "Contributions to the theory of testing statistical hypotheses. I. Unbiased critical regions of type A and type A_1," *Stat. Res. Mem.*, Vol. 1 (1936), pp. 1–37.

[2] J. Neyman, "Outline of a theory of statistical estimation based on the classical theory of probability," *Phil. Trans. Roy. Soc. London*, Series A, Vol. 236 (1937), pp. 333–380.

[3] F. N. David and J. Neyman, "Extension of the Markoff theorem on least squares," *Stat. Res. Mem.*, Vol. 2 (1938), pp. 105–116.

[4] G. W. Brown, "On small sample estimation," *Annals of Math. Stat.*, Vol. 18 (1947), pp. 582–585.

[5] A Wald, *Statistical Decision Functions*, John Wiley and Sons, 1950.
[6] G. Hunt and C. Stein, "Most stringent tests of statistical hypotheses," unpublished.
[7] E. L. Lehmann, "Some principles of the theory of testing hypotheses," *Annals of Math. Stat.*, Vol. 21 (1950), pp. 1-26.
[8] M. Peisakoff, "Transformation parameters," unpublished thesis, Princeton University, 1950.
[9] P. R. Halmos, "The theory of unbiased estimation," *Annals of Math. Stat.*, Vol. 17 (1946), pp. 34-43.
[10] P. L. Hsu, "Analysis of variance from the power function standpoint," *Biometrika*, Vol. 32 (1941), pp. 62-69.

ONE-SIDED CONFIDENCE CONTOURS FOR PROBABILITY DISTRIBUTION FUNCTIONS[1]

By Z. W. Birnbaum and Fred H. Tingey[2]

University of Washington

Summary. Let $F(x)$ be the continuous distribution function of a random variable X, and $F_n(x)$ the empirical distribution function determined by a sample $X_1, X_2, \cdots, X_n$. It is well known that the probability $P_n(\epsilon)$ of $F(x)$ being everywhere majorized by $F_n(x) + \epsilon$ is independent of $F(x)$. The present paper contains the derivation of an explicit expression for $P_n(\epsilon)$, and a tabulation of the 10%, 5%, 1%, and 0.1% points of $P_n(\epsilon)$ for $n = 5, 8, 10, 20, 40, 50$. For $n = 50$ these values agree closely with those obtained from an asymptotic expression due to N. Smirnov.

1. Introduction. Let X be a random variable with the continuous probability distribution function $F(x) = \text{Prob.}\,\{X \le x\}$. An ordered sample $X_1 \le X_2 \le \cdots \le X_n$ of X determines the empirical distribution function

$$F_n(x) = \begin{cases} 0 & \text{for } x < X_1, \\ \dfrac{k}{n} & \text{for } X_k \le x < X_{k+1}, \qquad k = 1, 2, \cdots, n-1, \\ 1 & \text{for } X_n \le x. \end{cases}$$

The function

$$F^+_{n,\epsilon}(x) = \min\,[F_n(x) + \epsilon, 1],$$

also determined by the sample, will be called an *upper confidence contour*. It is well known [2] that the probability

$$P_n(\epsilon) = \text{Prob.}\,\{F(x) \le F^+_{n,\epsilon}(x) \text{ for all } x\}$$

of $F(x)$ being everywhere majorized by $F^+_{n,\epsilon}(x)$ is independent of the distribution $F(x)$. An expression for $P_n(\epsilon)$ in determinant form was given by A. Wald and

[1] Presented to the American Mathematical Society on April 28, 1951.
[2] Research under the sponsorship of the Office of Naval Research.

J. Wolfowitz [2]. N. Smirnov [1] obtained the asymptotic expression

$$\lim_{n\to\infty} P_n\left(\frac{z}{\sqrt{n}}\right) = 1 - e^{-2z^2}. \tag{1.1}$$

The present paper contains the derivation of an explicit expression for $P_n(\epsilon)$, and a tabulation of values $\epsilon_{n,\alpha}$ such that

$$P_n(\epsilon_{n,\alpha}) = 1 - \alpha \tag{1.2}$$

for $\alpha = .10, .05, .01, .001$, and $n = 5, 8, 10, 20, 40, 50$. For $n = 50$ these values agree very closely with those obtained from Smirnov's asymptotic expression (1.1).

2. Two integral formulae. For any integer k, $1 \le k \le n$, we have

$$f_{k-1}(X_{k-1}) = \int_{X_{k-1}}^{1}\int_{X_k}^{1}\cdots\int_{X_{n-1}}^{1} dX_n \cdots dX_{k+1}\,dX_k = \frac{(1 - X_{k-1})^{n-k+1}}{(n-k+1)!} \tag{2.1}$$

This formula is well known and may be obtained by an easy induction.

For any integer $k \ge 0$ we have

$$\int_0^{\epsilon}\int_{X_1}^{(1/n)+\epsilon}\cdots\int_{X_k}^{(k/n)+\epsilon} dX_{k+1}\cdots dX_2\,dX_1 = \frac{\epsilon}{(k+1)!}\left(\epsilon + \frac{k+1}{n}\right)^k. \tag{2.2}$$

To prove (2.2) one shows by induction that the left-hand expression is equal to

$$\frac{\epsilon}{(m+2)!}\sum_{j=1}^{m+2}\binom{m+2}{j}\left(\epsilon + \frac{m+2-j}{n}\right)^{m+1}(-1)^{j-1},$$

which is equal to the right-hand term in view of the identity

$$\sum_{j=0}^{m+2}\binom{m+2}{j}\left(\epsilon + \frac{m+2-j}{n}\right)^{m+1}(-1)^{j-1} = 0.$$

3. An expression for $P_n(\epsilon)$.

THEOREM. *For $0 < \epsilon \le 1$ we have*

$$P_n(\epsilon) = 1 - \epsilon\sum_{j=0}^{[n(1-\epsilon)]}\binom{n}{j}\left(1 - \epsilon - \frac{j}{n}\right)^{n-j}\left(\epsilon + \frac{j}{n}\right)^{j-1}, \tag{3.0}$$

where $[n(1 - \epsilon)]$ = greatest integer contained in $n(1 - \epsilon)$.

PROOF. Since $P_n(\epsilon)$ does not depend on $F(x)$, we will assume that X has the probability distribution function

$$F(x) = \begin{cases} 0 & \text{for } x < 0, \\ x & \text{for } 0 \le x < 1, \\ 1 & \text{for } 1 \le x. \end{cases}$$

For this random variable, $P_n(\epsilon)$ is the probability that the ordered sample

$$0 \le X_1 \le X_2 \le \cdots \le X_n \le 1 \tag{3.1}$$

falls into the region

$$\begin{aligned} X_{j-1} \le X_j &\le \frac{j-1}{n} + \epsilon \qquad \text{for } j = 1, \cdots, K+1, \\ X_{j-1} \le X_j &\le 1 \qquad \text{for } j = K+2, \cdots, n, \end{aligned} \tag{3.2}$$

where $X_0 = 0$ and $K = [n(1 - \epsilon)]$. Since the probability density of an ordered sample $(X_1, X_2, \cdots, X_n)$ is equal to $n!$ in the region (3.1) and to zero elsewhere, the probability of (3.2) is equal to

$$P_n(\epsilon) = n!J(\epsilon, n, K), \tag{3.3}$$

where

$$J(\epsilon, n, K) = \int_0^{\epsilon} \int_{X_1}^{(1/n)+\epsilon} \int_{X_2}^{(2/n)+\epsilon} \cdots \int_{X_K}^{(K/n)+\epsilon} \int_{X_{K+1}}^{1} \int_{X_{K+2}}^{1} \cdots \int_{X_{n-1}}^{1} dX_n \cdots dX_{K+3}\, dX_{K+2}\, dX_{K+1} \cdots dX_3\, dX_2\, dX_1. \tag{3.4}$$

By (2.1) we see that

$$J(\epsilon, n, k) = \int_0^{\epsilon} \int_{X_1}^{(1/n)+\epsilon} \int_{X_2}^{(2/n)+\epsilon} \cdots \int_{X_k}^{(k/n)+\epsilon} \frac{(1 - X_{k+1})^{n-k-1}}{(n - k - 1)!}\, dX_{k+1} \cdots dX_3\, dX_2\, dX_1. \tag{3.5}$$

We will prove by induction

$$J(\epsilon, n, k + 1) = J(\epsilon, n, k) - \frac{\epsilon}{n!}\binom{n}{k+1}\left(1 - \epsilon - \frac{k+1}{n}\right)^{n-k-1} \cdot \left(\epsilon + \frac{k+1}{n}\right)^{k}, \tag{3.6}$$

for any integer $0 \leq k \leq n - 1$. For $k = 0$, (3.6) can be verified directly. Assuming (3.6) for $k \leq m$, we obtain

$$J(\epsilon, n, m + 1)$$
$$= \int_0^{\epsilon} \int_{X_1}^{(1/n)+\epsilon} \int_{X_m}^{(m/n)+\epsilon} \int_{X_{m+1}}^{((m+1)/n)+\epsilon} \frac{(1 - X_{m+2})^{n-m-2}}{(n - m - 2)!}\, dX_{m+2}\, dX_{m+1} \cdots dX_2\, dX_1$$
$$= \int_0^{\epsilon} \int_{X_1}^{(1/n)+\epsilon} \cdots \int_{X_m}^{(m/n)+\epsilon} \frac{(1 - X_{m+1})^{n-m-1}}{(n - m - 1)!}\, dX_{m+1} \cdots dX_2\, dX_1$$
$$- \frac{\left(1 - \epsilon - \frac{m+1}{n}\right)^{n-m-1}}{(n - m - 1)!} \int_0^{\epsilon} \int_{X_1}^{(1/n)+\epsilon} \cdots \int_{X_m}^{(m/n)+\epsilon} dX_{m+1} \cdots dX_2\, dX_1,$$

and, by the assumption of induction and (2.2), this is

$$J(\epsilon, n, m) - \frac{\epsilon}{n!}\binom{n}{m+1}\left(1 - \epsilon - \frac{m+1}{n}\right)^{n-m-1}\left(\epsilon + \frac{m+1}{n}\right)^{m},$$

which proves (3.6).

Noting that $J(\epsilon, n, 0) = \frac{1}{n!}[1 - (1 - \epsilon)^n]$, one obtains from (3.6)

$$J(\epsilon, n, k) = \frac{1}{n!}[1 - (1 - \epsilon)^n] - \frac{\epsilon}{n!}\sum_{j=1}^{k}\binom{n}{j}\left(1 - \epsilon - \frac{j}{n}\right)^{n-j}\left(\epsilon + \frac{j}{n}\right)^{j-1}.$$

This, together with (3.3) completes the proof of (3.0).

Remark. Setting $F^{-}_{n,\epsilon}(x) = \max[F_n(x) - \epsilon, 0]$, one easily verifies that Prob. $\{F(x) \geq F^{-}_{n,\epsilon}(x)$ for all $x\}$ is equal to $P_n(\epsilon)$, and hence also is given by (3.0).

4. Tabulation of $\epsilon_{n,\alpha}$ and comparison with asymptotic values. Table 1 contains numerical solutions $\epsilon_{n,\alpha}$ of equation (1.2), computed to a number of digits sufficient to assure that $|P_n(\epsilon_{n,\alpha}) - (1 - \alpha)| < 5\cdot 10^{-5}$.

TABLE 1.[3]

Solutions $\epsilon_{n,\alpha}$ of equation (1.2)

n \ α	.100	.050	.010	.001
5	.4470	.5094	.6271	.7480
8	.3583	.4096	.5065	.6130
10	.3226	.3687	.4566	.5550
20	.23155	.26473	.3285	.4018
40	.16547	.18913	.2350	.2877
50	.14840	.16959	.2107	.2581

Setting $z/\sqrt{n} = \bar{\epsilon}_{n,\alpha}$ in (1.1), one obtains for large n the asymptotic values

$$\bar{\epsilon}_{n,\alpha} = \sqrt{\frac{1}{2n}\log\frac{1}{\alpha}}. \tag{4.1}$$

These values are presented in Table 2.

TABLE 2

Values of $\bar{\epsilon}_{n,\alpha} = \sqrt{\frac{1}{2n}\log\frac{1}{\alpha}}$

n \ α	.100	.050	.010	.001
5	.4799	.5473	.6786	.8311
8	.3794	.4327	.5365	.6571
10	.3393	.3870	.4799	.5877
20	.2399	.2737	.3393	.4156
40	.1697	.1935	.2399	.2938
50	.1517	.1731	.2146	.2628

A comparison of the two tables indicates that, for the probability levels $.001 \leq \alpha \leq .1$, the asymptotic values $\bar{\epsilon}_{n\alpha}$, are greater than the "exact" values

[3] The authors wish to express their appreciation to the National Bureau of Standards, Institute for Numerical Analysis, for performing the computations which are summarized in this table.

$\epsilon_{n,\alpha}$ so that the error committed by using $\bar{\epsilon}_{n,\alpha}$ instead of $\epsilon_{n,\alpha}$ would be in the safe direction, and that this error becomes already very small for $n = 50$.

REFERENCES

[1] N. Smirnov, "Sur les écarts de la courbe de distribution empirique," *Rec. Math.* (*Mat. Sbornik*), N. S. Vol. 6 (48) (1939), pp. 3–26.

[2] A. Wald and J. Wolfowitz, "Confidence limits for continuous distribution functions," *Annals of Math. Stat.*, Vol. 10 (1939), pp. 105–118.

ON THE ESTIMATION OF CENTRAL INTERVALS WHICH CONTAIN ASSIGNED PROPORTIONS OF A NORMAL UNIVARIATE POPULATION

By G. E. Albert and Ralph B. Johnson

University of Tennessee and Clemson Agricultural College

Summary. For samples of any given size $N \geq 2$ from a normal population, Wilks [1] has shown how to choose the parameter λ_p so that the expected coverage of the interval $\bar{x} \pm \lambda_p s$ will be $1 - p$. The present paper treats the choice of the minimal sample size N necessary to effect a certain type of statistical control on the fluctuation of that coverage about its expected value; a brief table of such minimal sample sizes is given.

1. Introduction. Let $F(y)$ denote the normal cumulative distribution function

$$F(y) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{y} e^{-(u-m)^2/(2\sigma^2)}\, du. \tag{1}$$

If p is any number in the range $0 < p < 1$, factors $\lambda(p)$ are well known such that the proportion

$$A = F(m + \lambda\sigma) - F(m - \lambda\sigma) \tag{2}$$

of the probability between $m \pm \lambda\sigma$ will equal $1 - p$.

If m and σ are unknown, it is natural to consider the random variable

$$A(\bar{y}, s; \lambda) = F(\bar{y} + \lambda s) - F(\bar{y} - \lambda s), \tag{3}$$

where $\bar{y} = \sum_{n=1}^{N} y_n/N$ and $s = \left\{\sum_{i=1}^{N} (y_i - \bar{y})^2/(N-1)\right\}^{\frac{1}{2}}$.

Obviously λ cannot be chosen to guarantee $A(\bar{y}, s; \lambda) = 1 - p$. S. S. Wilks [1] has shown that, for a random sample of size N, the expectation of (3) is $1 - p$,

$$EA(\bar{y}, s; \lambda) = 1 - p, \tag{4}$$

if the parameter λ is chosen as

$$\lambda = t_p \sqrt{\frac{N+1}{N}}. \tag{5}$$

In (5) t_p is such that for Student's t-distribution of $N - 1$ degrees of freedom

$$Pr[\,|\,t\,|\, \geq t_p] = p.$$

Wilks' study of the variability of $A(\bar{y}, s; \lambda)$ was based upon an approximate consideration of the variance of A. It is the purpose of this paper to present more precise results in this latter connection.

Let d_1, d_2 and α be assigned positive numbers satisfying the inequalities $0 \leq 1 - p - d_1 < 1 - p + d_2 \leq 1$, and $0 < \alpha < 1$. It is shown that if λ be chosen as in (5), the requirement

$$Pr[1 - p - d_1 \leq A(\bar{y}, s; \lambda) \leq 1 - p + d_2] \geq \alpha \tag{6}$$

places a lower bound on the sample size N. It is clear that if d_1 and d_2 are small and α near unity, (6) places a control on the variability of $A(\bar{y}, s; \lambda)$ about its expectation $1 - p$.

TABLE I

Smallest N for which (6) holds

p		.01		.05			.25			.50		
α		.95	.99	.80	.95	.99	.80	.95	.99	.80	.95	.99
d_1	d_2											
.075	.05	—	—	—	24	49	54	128	226	44	108	197
.05	.05	—	—	—	43	92	76	174	298	63	144	243
.025	.025	—	—	65	159	299	298	692	1194	245	567	975
.035	.015	—	—	107	274	510	420	1332	2628	337	1079	2184
.05	.01	12	27	196	640	1230	813	2991	5983	649	2488	4928
.025	.01	26	64	226	641	1230	907	2993	5983	725	2487	4928
.02	.01	37	88	254	657	1231	1025	3015	5982	825	2502	4928
.01	.01	110	319	428	1009	1750	1846	4319	7456	1507	3540	6084

Methods devised by Wald and Wolfowitz [2] are easily adapted to the approximate calculation of the probability (6).

Table I presents *minimal* values of the sample size N to effect the control (6) for various values of the constants p, d_1, d_2 and α. The indication is clear that the *prediction of probability intervals based upon the estimates $\bar{y}$ and s from small samples is not very reliable.*

2. The expectation of A and the probability (6). Writing $u = (\bar{y} - m)/\sigma$ and $v = s/\sigma$, $A(\bar{y}, s; \lambda)$ becomes

$$A^*(u, v; \lambda) = \frac{1}{\sqrt{2\pi}} \int_{u-\lambda v}^{u+\lambda v} e^{-\frac{1}{2}t^2}\, dt. \tag{7}$$

It is well known that the variables $u\sqrt{N}$ and $(N - 1)v^2$ are independently distributed, the first being normal with zero mean and unit variance and the second

being chi-square with $N - 1$ degrees of freedom. One readily derives (Wilks [1])

$$E(A) = Pr\left[\,|t| \leq \lambda\sqrt{\frac{N}{N+1}}\,\right],$$

where t has Student's distribution with $N - 1$ degrees of freedom. Setting this equal to $1 - p$, the choice (5) for λ is obtained.

To calculate the probability (6), one integrates the joint frequency function $f(u, v)$ over that portion of the half plane $-\infty < u < \infty, v > 0$ on which $1 - p - d_1 \leq A^* \leq 1 - p + d_2$. To perform the integration, one proceeds as in Wald and Wolfowitz [2] where a similar problem is solved. Define two functions

$$v_r = v_r(u), \qquad r = 1, 2 \tag{8}$$

by the equations

$$A^*(u, v_r; \lambda) = 1 - p + (-1)^r d_r, \qquad r = 1, 2, \tag{9}$$

where A^* is defined by (7) and λ is given by (5). The functions $v_r(u)$ are monotone increasing relative to $|u|$. It follows that

$$Pr\{1 - p - d_1 \leq A(\bar{y}, s; \lambda) \leq 1 - p + d_2\} = \sqrt{\frac{N}{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}Nu^2}P(u)\,du, \tag{10}$$

where

$$P(u) = Pr\{(N-1)v_1^2(u) < \chi^2 < (N-1)v_2^2(u)\}, \tag{11}$$

χ^2 being distributed as chi-square with $N - 1$ degrees of freedom.

The formulas (10) and (11) are too unwieldy for much computation. Following Wald and Wolfowitz [2] again, one can show that a good approximation for large N is

$$Pr\{1 - p - d_1 \leq A(\bar{y}, s; \lambda) \leq 1 - p + d_2\} \cong P(N^{-\frac{1}{2}}), \tag{12}$$

the right member being given by (11).

3. Computational procedure. For a given set of values of p, d_1, and d_2, one may now tabulate (12) against N by the following steps. Using λ as given by (5), the $v_r = v_r(N^{-\frac{1}{2}})$ defined by (9) are found by trial and error from a standard normal distribution table. Then (11) and (12) give the control probability. One easily picks out the minimal N for which (6) is satisfied. Tables of the incomplete gamma function [3] are available and the authors are in possession of graphs of the chi-square distribution prepared from these tables by the use of spline curves. The detail of the graphs is sufficient for three-decimal accuracy in reading probabilities. For small values of N and values beyond the range of tables, a variety of standard methods of approximation for (11) were used.

Lower and upper bounds for the interval (10) are easily devised using obvious approximate quadrature methods. See Wald and Wolfowitz [2] in this connection. The small values of N in Table I were checked by such a device. The

authors are confident that the computation was sufficiently accurate to make the table useful for practical purposes.

4. Generalization. The formulation of the problem discussed above may be generalized to the case in which the mean m of the distribution (1) depends linearly upon k sure variables $x_1, x_2, \cdots, x_k$. The N observations are then N $(k + 1)$-tuples $(y_i; x_{i1}, x_{i2}, \cdots, x_{ik})$, $i = 1, 2, \cdots, N$, and the mean has the form

$$m = \alpha + \sum_{j=1}^{k} \beta_j(X_j - \bar{x}_j)$$

for an arbitrary set of values $X_1, X_2, \cdots, X_k$ of the sure variables. Referring to Cramér ([4], pages 551 and 552) for notations and formulas in order to save space here, one replaces the interval estimate $(\bar{y} \pm \lambda s)$ above by the interval from R_1 to R_2, where

$$R_r = \alpha^* + \sum_{j=1}^{k} \beta_j^*(X_j - \bar{x}_j) + (-1)^r \lambda^* \sigma^*, \qquad r = 1, 2,$$

$$\lambda^* = t_p \sqrt{\frac{N + M}{N - k - 1}},$$

and

$$M = 1 + \sum_{i,j=1}^{k} \frac{L_{ij}}{L} (X_i - \bar{x}_i)(X_j - \bar{x}_j).$$

Here t_p is chosen as in (5) except that the degrees of freedom are now $N - k - 1$.

For this generalization, when N/M is large, the control probability (6) is approximated by $P(M^{\frac{1}{2}}/N^{\frac{1}{2}})$ where $P(u)$ is given by (11). Organized computation for this generalization does not seem feasible since the values of the quadratic form M may vary greatly from one application to another.

REFERENCES

[1] S. S. Wilks, "Determination of sample sizes for setting tolerance limits," *Annals of Math. Stat.*, Vol. 12 (1941), pp. 91–96.

[2] A. Wald and J. Wolfowitz, "Tolerance limits for a normal distribution," *Annals of Math. Stat.*, Vol. 17 (1946), pp. 208–215.

[3] K. Pearson, *Tables of the Incomplete Gamma Function*, Cambridge University Press 1922.

[4] H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, 1946.

ON DEPENDENT TESTS OF SIGNIFICANCE IN THE ANALYSIS OF VARIANCE[1]

By A. W. Kimball
Oak Ridge National Laboratory

1. Introduction. Some statisticians and other practitioners of the analysis of variance have expressed concern over the fact that many experimental designs lead to multiple tests of significance which are not independent in the probability sense. Factorials, latin squares, lattices, etc. have the advantage of enabling a research worker to test several hypotheses in one experiment, but all tests ordinarily depend on the same estimate of population variance. It is argued that whatever error is present in this estimate for a particular experiment will affect all tests of hypothesis in the same manner, and one tends either to accept or reject a large proportion of the hypotheses when the population variance is respectively overestimated or underestimated. The difficulty can be avoided by performing a separate experiment for each hypothesis to be tested, but this would contradict the whole philosophy of experimental design.

This paper deals with an attempt to evaluate the effect of dependency among the tests of significance when each experiment is treated as a unit regardless of the number of hypotheses tested per experiment. From this point of view if all null hypotheses are true, an error is committed if one or more of the hypotheses are rejected. It is shown that the probability of making no errors of the first kind in one experiment is greater when the tests are dependent than when they are independent. For those who prefer this way of looking at the problem, the doubts expressed in the first paragraph should be dispelled. The situation in which risks are calculated using the hypothesis rather than the experiment as a unit is not considered.

In the following sections it is assumed that samples are taken independently from normal populations having the same variance and having means additively related in a manner defined by the design of the experiment. These are the usual assumptions associated with analysis of variance models in which the parameters are population means (as distinguished from components of variance models).

2. Case of two dependent tests of hypothesis. We shall consider first the case of an analysis of variance in which two hypotheses are tested using the same error variance for each test. A well known example of this case occurs in the analysis of variance with two criteria of classification where the effects of both rows and columns are to be tested. In the usual cases, formulation as a general linear hypothesis leads to three quadratic forms, q_1, q_2, and q_3, which are independently distributed as χ^2 with n_1, n_2, and n_3 degrees of freedom, respectively.[2] The likelihood ratio statistics for testing the two hypotheses are then

$$F_1 = \frac{q_1/n_1}{q_3/n_3} \quad \text{and} \quad F_2 = \frac{q_2/n_2}{q_3/n_3}.$$

[1] This work was begun while the author was at the USAF School of Aviation Medicine, Randolph Field, Texas.

[2] For a more complete statement, see [1], p. 177.

If the critical region for the rejection of each null hypothesis is of size α, the probability of making no errors of the first kind is given by

$$P\{F_1 \leq F_{1\alpha}, F_2 \leq F_{2\alpha}\},$$

where $F_{1\alpha}$ and $F_{2\alpha}$ are the 100α per cent points of the distributions of F_1 and F_2, respectively. We shall prove[3] that

(1) $$P\{F_1 \leq F_{1\alpha}, F_2 \leq F_{2\alpha}\} > P\{F_1 \leq F_{1\alpha}\} \cdot P\{F_2 \leq F_{2\alpha}\}.$$

Since q_1, q_2, and q_3 are independent, their joint density is the product of three χ^2 densities. Clearly (1) may be written

(2) $$P\{q_1 \leq k_1 q_3, q_2 \leq k_2 q_3\} > P\{q_1 \leq k_1 q_3\} \cdot P\{q_2 \leq k_2 q_3\},$$

where $k_1 = n_1 F_{1\alpha}/n_3$, $k_2 = n_2 F_{2\alpha}/n_3$. Expressed in integral form, (2) become

(3) $$\int_0^\infty f_1(q_3) f_2(q_3) f_3(q_3)\, dq_3 > \int_0^\infty f_1(q_3) f_3(q_3)\, dq_3 \int_0^\infty f_2(q_3) f_3(q_3)\, dq_3,$$

where for $i = 1$ or 2, $f_i(q_3)$ is the integral from zero to $k_i q_3$ of a χ^2 density with n_i degrees of freedom, while $f_3(q_3)$ is the χ^2 density function with n_3 degrees of freedom. Since $f_1(q_3)$ and $f_2(q_3)$ are positive strictly monotonically increasing functions of q_3, and $f_3(q_3)$ is a density function, (3) may be written

(4) $$E[f_1(q_3) f_2(q_3)] > E[f_1(q_3)] \cdot E[f_2(q_3)],$$

where the expected values are taken over the probability distribution of χ^2.

The inequality expressed in (4) may be proved as a special case of the following theorem.[4]

THEOREM. *If $f(x) \geq 0$ and $g(x) \geq 0$ are both strictly monotonically increasing functions of a random variable x having the probability density $h(x)$ $(0 \leq x \leq \infty)$, and if both $f(x)$ and $g(x)$ have finite expectations, then*

$$E[f(x)g(x)] - E[f(x)] \cdot E[g(x)] > 0.$$

PROOF. We may write

$$E[f(x)\, g(x)] - E[f(x)] \cdot E[g(x)] = \int_0^\infty f(x)\{g(x) - E[g(x)]\} h(x)\, dx$$
$$= I,$$

say. Because of the monotonicity of $g(x)$, there must exist a quantity $x_0 > 0$ such that $g(x_0) = E[g(x)]$. It follows that

$$I = -\int^{x_0} f(x)\{E[g(x)] - g(x)\} h(x)\, dx$$
$$+ \int_{x_0}^\infty f(x)\{g(x) - E[g(x)]\} h(x)\, dx$$
$$= -I_1 + I_2,$$

[3] The trivial cases in which either $F_{1\alpha}$ or $F_{2\alpha}$ or both are either zero or infinite are excluded.

[4] The author is indebted to Dr. Max Halperin for the proof of this theorem.

say. Since

$$\int_0^\infty \{g(x) - E[g(x)]\}h(x)\,dx = 0,$$

we must have

$$-\int_0^{x_0} \{g(x) - E\,[g(x)]\}h(x)\,dx = \int_{x_0}^\infty \{g(x) - E[g(x)]\}h(x)\,dx$$
$$= J,$$

say. Furthermore, since $f(x)$ is a strictly monotonically increasing function of x, it follows that

$$I_1 < f(x_0)J, \qquad I_2 > f(x_0)J.$$

Therefore, $I_2 - I_1 = I > 0$, and the theorem is proved.

It is obvious that the foregoing theorem may be applied directly to prove the validity of (4). This in turn verifies (1).

Although the proof in this section was introduced by reference to a specific model in the analysis of variance, it is clearly valid for any two F-tests of significance which satisfy the relationships with respect to q_1, q_2, and q_3, and in general for any n_1, n_2, and n_3.

3. Extension to several dependent tests of hypothesis. The extension of (1) to more than two tests of significance is straightforward. If there are three F-tests, we must show that

$$(5) \quad \int_0^\infty f_0(q_3)\, f_1(q_3)\, f_2(q_3)\, f_3(q_3)\, dq_3$$
$$> \int_0^\infty f_0(q_3)\, f_3(q_3)\, dq_3 \int_0^\infty f_1(q_3)\, f_3(q_3)\, dq_3 \int_0^\infty f_2(q_3)\, f_3(q_3)\, dq_3,$$

where $f_0(q_3)$ is a function similar to $f_1(q_3)$ and $f_2(q_3)$ resulting from the third test of significance. From Section 2 we know that

$$(6) \quad \int_0^\infty f_0(q_3)\, f_1(q_3)\, f_2(q_3)\, f_3(q_3)\, dq_3 > \int_0^\infty f_0(q_3)\, f_3(q_3)\, dq_3 \int_0^\infty f_1(q_3)\, f_2(q_3)\, f_3(q_3)\, dq_3,$$

since $f_0(q_3)$ and $f_1(q_3)f_2(q_3)$ satisfy the requirements of the theorem. But from (3) we may make an obvious substitution in the right-hand side of (6) which reduces it to (5). Clearly this simple procedure may be repeated as often as necessary to prove the extension of (1) to any number of F-tests of significance in which the numerators of the test statistics are mutually independent, and each is independent of the denominator which is the same for all statistics.

The author wishes to thank Professors J. W. Tukey and H. Levene for helpful suggestions in the preparation of this manuscript.

REFERENCE

[1] S. S. Wilks, *Mathematical Statistics*, Princeton University Press, 1946.

ON A CONNECTION BETWEEN CONFIDENCE AND TOLERANCE INTERVALS

By Gottfried E. Noether

Boston University

The purpose of this note is to point out the close connection which exists between confidence intervals for the parameter p of a binomial distribution and tolerance intervals.

Let k be the number of successes in a random sample of size n from a binomial population with probability p of success in a single trial. Then it is well known that a confidence interval with confidence coefficient at least $1 - \alpha_1 - \alpha_2$ for the parameter p is given by

$$p_1(k) < p < p_2(k), \tag{1}$$

where $p_1(k)$ and $p_2(k)$ are determined by $I_{p_1(k)}(k, n - k + 1) = \alpha_1$ and $I_{1-p_2(k)}(n - k, k + 1) = 1 - I_{p_2(k)}(k + 1, n - k) = \alpha_2$, respectively, $I_x(a, b) = [\Gamma(a + b)/(\Gamma(a)\Gamma(b))] \int_0^x u^{a-1}(1 - u)^{b-1}\, du$ being the incomplete B-function.

Let $X_1, \cdots, X_n$ represent a random sample of size n from a population having continuous cdf $F(x)$. For simplicity assume that the X's are already arranged in increasing order of size and define $X_0 = -\infty$, $X_{n+1} = +\infty$. The coverage provided by the interval (X_i, X_{i+1}), $i = 0, 1, \cdots, n$, is called an elementary coverage.[1] If we then let U_r stand for the sum of r elementary coverages, $U_r > U_r(\alpha)$ unless an event of probability α has occurred, where $U_r(\alpha)$ is defined by $\alpha = [\Gamma(n + 1)/(\Gamma(r)\Gamma(n - r + 1))] \int_0^{U_r(\alpha)} u^{r-1}(1 - u)^{n-r}\, du = I_{U_r(\alpha)}(r, n - r + 1)$.

In this notation (1) becomes

$$U_k(\alpha_1) < p < U_{k+1}(1 - \alpha_2).$$

Thus the lower end point of a confidence interval for p on the basis of k observed successes is determined by the corresponding lower limit for the sum of k elementary coverages, while the upper end point is determined by the corresponding upper limit of the sum of $(k + 1)$ elementary coverages. The reason for this becomes obvious if we look at the k successes as the observations $X_1, \cdots, X_k$ which are smaller than the p-quantile q_p of $F(x)$, so that the coverage U_k of the chance interval (X_0, X_k) provides an "inner" estimate of p, while the coverage U_{k+1} of the chance interval (X_0, X_{k+1}) provides an "outer" estimate.

We may ask what kind of a confidence interval we obtain if we consider as successes the k observations belonging to an arbitrary interval I for which $\int_I dF(x) = p$, as long as I does not coincide with either $(-\infty, q_p)$ or $(q_{1-p}, +\infty)$.

[1] For rigorous definitions and formulas see, e.g., Wilks [1], p. 13.

It is easily seen that an "outer" estimate of p is still given by U_{k+1}. However, an "inner" estimate is now given by U_{k-1}, leading to a lower end point of the confidence interval which is unnecessarily small.

The method of obtaining a confidence interval for p discussed in this note is in a certain sense the reverse of the method discussed in an earlier paper of the author [2]. There it was shown how confidence intervals for p can be used to obtain confidence intervals for quantiles, which then can be used to obtain tolerance intervals.

REFERENCES

[1] S. S. Wilks, "Order statistics," *Bull. Am. Math. Soc.*, Vol. 54 (1948), pp. 6–50.

[2] G. E. Noether, "On confidence limits for quantiles," *Annals of Math. Stat.*, Vol. 19 (1948), pp. 416–419.

ABSTRACTS OF PAPERS

(Abstracts of papers presented at the Minneapolis meeting of the Institute, September 4–7, 1951)

1. On Stieltjes Integral Equations of Stochastic Processes. Maria Castellani, University of Kansas City.

This paper considers two methods of solving certain S-integral equations.

a. *A Fredholm-Stieltjes integral equation of generating functions.* We give the F-S integral equation $\int_E A(s, x)\, dg(x) = f(s)$, where $A(s, x) = \Sigma_{k=0}^{\infty} \alpha_k(x)s^{-k}$ and $f(s) = \Sigma a_k s^k$ for $s \to \varphi(s)$ and $a_0 = 0$ if $k = 0$. Let us assume that $u(x)$ and $v(x)$ are respectively solutions of $\int_E A(s, x)\cdot A(-s_1, x)\, du(x) = 1/(S - S_1)$ and $\int_E A(s, x)\, dv(x) = 0$. If we consider

$$\int_E A(s, x)A(-s_1, x)f(s_1)\, du(x) = f(s_1)/(S - S_1)$$

and if $\gamma(x)$ is the coefficient of $-1/S_1$ in the serial expansion of $A(-s, x)f(s_1)$, then under fairly general conditions the required solutions are given, almost everywhere, by $g(x) = \text{const.} \int^x dv(x) + \int^x \gamma(x)\, du(x)$. The proof is based on a Murphy D'Arcais linear operator and on the ρ operation of S-integrals.

b. *A Volterra-Stieltjes integral of recurrent random functions.* Let us have over a time interval (τ, t) an unknown rf$\delta(t - \tau)$ satisfying the following recursive equation: $\delta(t - \tau) = \delta(\tau) - \int_\tau^t \delta(x - \tau)\rho(x)\, dF(x)$ where $F(x)$ is a df and $\rho(x)$ is bounded. We assume the interval divided into n parts and also that the set of the n discrete values of δ satisfy the following relation: $\delta(t - \tau)/\delta(\tau) = \Pi_{s=\tau}^{s=t}(1 - \rho(s)\Delta F(s))$. If $F = F_1 + F_2$, where the F_1 is a continuous function and F_2 is a jump function over a set S of points, then by a generalized method of Cantelli, taking finer and finer partitions, we obtain as a limit $\delta(t - \tau)/\delta(\tau) = \left[\exp\left(-\int_\tau^t \rho(x)\, dF_1(x)\right)\right] \Pi_{s \epsilon S}(1 - \rho(s)\, dF_2(s))$. This gives almost everywhere the required solutions.

2. **An Unfavorable Aspect of the Liklihood Ratio Test.** L. M. COURT, Rutgers University.

The likelihood ratio test has many desirable properties. For example, it is not only consistent, but as Wald has shown, uniformly consistent. Still, it can at times be a poor test, e.g., under certain circumstances when the size of the test region is properly selected, the probability of rejecting the hypothesis to be tested when it is true *exceeds* the probability of rejecting it when any alternative is true. Both Stein and Rubin have given examples of this. Stein's example, quoted by Lehmann in his notes on "Testing Hypotheses" (pp. 1–5) consists of a family of *discrete* distributions (five-valued, to be precise) in which a simple hypothesis is tested against a composite alternative. The writer, using a geometrical construction, gives an example (actually, a broad class of examples, much broader than Stein's which is a 2-parameter class) in which the distributions are *continuous*. The hypothesis to be tested is first simple, then composite; the alternative, always composite.

3. **Impartial Decision Rules and Sufficient Statistics.** RAGHU RAJ BAHADUR AND LEO A. GOODMAN, University of Chicago.

In the following, (1) refers to the paper "On a problem in the theory of k populations," by R. R. Bahadur (*Annals of Math. Stat.* Vol. 21 (1950), pp. 362–375). The present paper provides certain improvements of the main result contained in (1). The authors define the class of impartial decision rules in terms of permutations of the k samples (rather than in terms of the k ordered values of an arbitrarily chosen real valued statistic (cf. (1)). This definition is intuitively more appealing than the one adopted in (1), and permits a unified treatment of discrete and absolutely continuous populations. The authors show that if the same function is a sufficient statistic for each of k independent samples of equal size, then the conditional expectation given the sufficient statistics of an impartial decision rule is also an impartial decision rule. They also give a characterization of impartiality which relates the present definition to that of (1). These results, together with Theorem 1 of (1), yield the desired improvements. An illustration of the argument indicated here is given.

4. **Contributions to the Statistical Theory of Counter Data.** G. E. ALBERT, University of Tennessee and Oak Ridge National Laboratory, AND M. L. NELSON, Oak Ridge National Laboratory.

Let a sequence f of events be such that the number occurring in time T is a chance variable having a Poisson distribution with mean aT, $a > 0$. A counting device generates a new sequence g since, due to a resolving time $u > 0$, it fails to record all of f. An event in f (i) *can* be recorded only if none has been recorded during a time u preceding it, (ii) *will* be recorded if it follows its predecessor in f by more than time u, (iii) either will be recorded with probability p or not recorded with probability $1 - p$ if it follows its predecessor in f by time $\leq u$. The choices $p = 0$ and $p = 1$ give the so called Type I and Type II counters respectively. The distribution theory of g is obtained as a generalization of the Type II theory given by Feller in "On probability problems in the theory of counters," *Courant Anniversary Volume*, 1948. The Cornish-Fisher normalization is applied to obtain confidence intervals for the constant a from observations on g of either the time to a preassigned count or the count to a preassigned time. These intervals turn out essentially independent of p whenever the product au is small; thus the Type I theory reported at an earlier meeting covers most of the cases of practical importance.

5. **On the Use of Wald's Classification Statistic.** HARMAN LEON HARTER, Michigan State College.

In 1944 Wald published a paper introducing the statistic V and giving a general outline of its use in problems of classification. Recently the author published a paper giving the distribution of V in various cases. The present paper takes the form of an effort to relate the two earlier papers and apply the results of the latter. The technique of classifying an individual into one of two groups is studied in detail. Let the individual under consideration belong to the population π, which is known to be identical with one or the other of two known populations π_1 and π_2. Then one may wish to test the hypothesis $H_1: \pi = \pi_1$, against the alternative hypothesis $H_2: \pi = \pi_2$. The values of the statistic V under these two hypotheses are given, and a method is outlined for testing H_1 against H_2, where an error of the second kind is k times as costly as an error of the first kind. A numerical example is given for the univariate case, for which the distribution of V is given in the author's earlier paper. The same procedure can be applied in the multivariate case when the distribution is known.

6. **Polynomial Determination in a Field of Integers Modulo P.** EDWARD C. VARNUM, Barber-Colman Company.

From a study of integers mod 2 applied to on-off relay circuits, a generalization to any prime modulus, p, has been made to construct a $p^n \times p^n$ matrix by which a polynomial in n variables may be determined when the p^n values of the polynomial are given for all the combinations of the p values of each of the n variables.

7. **About Some Symmetrical Distributions from the Perks' Family of Functions.** JOSEPH TALACKO, Marquette University.

The Perks' system of functions includes a family of symmetrical nonnormal distributions, from which two probability densities are of growing interest in theoretical statistics: the Verhulst's distribution (logistic distribution) and the hyperbolic cosine distribution. In the first part of this paper properties of this family of probability functions are discussed and the characteristic functions for Verhulst's and the hyperbolic secant distributions introduced. The Verhulst's probability function $f(t) = \delta e^{-\delta t}/(1 + e^{-\delta t})^2$ has $C(\nu) = (\pi\nu/\delta)$ cosech $(\pi\nu/\delta)$, and the hyperbolic secant probability function $\varphi(t) = (2\delta/\pi)(1/(e^{\delta t} + e^{-\delta t}))$ has $C(\nu) =$ sech $(\pi\nu/(2\delta))$. The second part is concerned with some previously uninvestigated distributions of certain statistics for samples from Verhulst's population. In particular, distributions of sample means and sum of squares are discussed. In an appendix a table of numerical values of Verhulst's functions is given.

8. **A Large-Sample Test for the Variation of Sample Covariance Matrices.** DAYLE D. RIPPE, University of Michigan.

A test criterion is developed to determine whether a given sample covariance matrix could be obtained as a result of taking a random sample of size N from a k-variate normal population with a given covariance matrix. The test is based upon the fact that the maximum likelihood estimate of the population covariance u_{ij} is $\hat{u}_{ij} = m_{ij} = (N - 1)^{-1} \sum_{\alpha=1}^{N} (x_{i\alpha} - \bar{x}_i)(x_{j\alpha} - \bar{x}_j)$. The test criterion for large samples is $\lambda = (N - 1)(\ln | u_{ij} | - \ln | m_{ij} | + \sum_i \sum_j u^{ij} m_{ij} - k)$, where λ is distributed as a chi-square with $\frac{1}{2}k(k + 1)$ degrees of freedom minus the number of independent linear restrictions among the variables $(m_{ij} - u_{ij})$. The results of the application of this criterion to the sampling problems in correlation theory compare favorably with exact sampling results, and the range of application is extended considerably. The criterion is sufficiently general in application to furnish a large-sample test for completeness of factorization in matrix factorization (or factor analysis) for the case of complete initial estimates of the communalities. It is applicable to any of the common orthogonal forms of solution. The

degrees of freedom of the chi-square involved is $\frac{1}{2}(k - s)(k - s + 1)$ after s of the total of k possible common components have been removed. The test may also be applied to determine the significance of component loadings in the common factor solution.

9. **Probability Models for Analyzing Time Changes in Attitudes.** T. W. ANDERSON, Columbia University.

Statistical inference in Markov chains is studied with particular application to data in which the finite number of states are states of attitudes of individuals in "panel surveys." Each individual's sequence of states over a finite number of time points is considered as an observation from a Markov chain with the same stationary transition probability matrix $P = (p_{ij})$. $n_i(0)$ individuals hold state i at the time origin. The maximum likelihood estimate of P is obtained. Tests are obtained for the hypothesis that P is a given matrix (or that certain elements are given numbers) and for the hypothesis that the transition matrix is stationary against alternatives that it varies over time. Extension of results to higher order cases is straightforward. A test of the hypothesis that the process is first order against the alternative that it is second order is given. When the state is defined in terms of two attitudes, a test is given for the hypothesis that the two attitudes change independently of each other. Asymptotic distributions of the estimates and of the test criteria are obtained under the assumption that $n_i(0) \to \infty$.

10. **The Variance of a Weighted Average Using Estimated Weights.** PAUL MEIER, Princeton University.

In various experimental designs (e.g., lattices) the problem of estimation involves averaging of two or more means with different variances. The proper weights (invariances) must be estimated from the experimental data. This problem has been treated by Cochran for the case of a large number of samples of equal size. We consider the case of two or more samples and find adjustments of order $O(\Sigma 1/n_i)$, both for the increase in variance and the bias of estimating the variance. Bounds on the increase in variance due to the use of estimated weights are given. Exact computations are made for several special cases. For the case of two means with weights based on ten degrees of freedom the adjustments reduce the maximum bias from approximately 15 per cent to less than 2 per cent.

11. **Distribution of Ratios of Quadratic Forms.** JOHN GURLAND, University of Chicago.

The problem of finding the distribution of a quadratic form and of a ratio of quadratic forms in normally distributed random variables is considered. By transforming the problem of ratios into one of linear combinations of independent variables each having a χ^2 distribution, a solution is given in terms of Laguerre polynomials which is more general than that of Pitman and Robbins ("Application of the method of mixtures to quadratic forms in normal variates," *Annals of Math. Stat.*, Vol. 20, 1949, pp. 552–560). The convergence of the expansion is established, and a new system of polynomials is suggested which would afford a solution for all distributions of quadratic forms and ratios of quadratic forms in normally distributed random variables. Once the convergence of expansions in terms of the new system polynomials is established, the system will be applicable in a much wider class of distributions than the Gram-Charlier series.

12. **The Large-Sample Power of Tests Based on Permutations of Observations. (Preliminary Report.)** WASSILY HOEFFDING, University of North Carolina.

The results of this paper can be illustrated by the following example. Let $t(x) = t(x_1, \cdots, x_N)$ be the usual t-statistic for testing whether two samples $x_1, \cdots, x_m$ and

$x_{m+1}, \cdots, x_N$ came from the same (normal) population. The critical region of the standard test of size α is of the form $|t(x)| \geq \lambda_N$. As $N \to \infty$, λ_N approaches a value $\lambda = \lambda(\alpha)$. A nonparametric test proposed by E. J. G. Pitman can be described as follows. Let $x^{(1)}, \cdots, x^{(N!)}$ be the $N!$ permutations of the sample values $x = (x_1, \cdots, x_N)$, so numbered that $|t(x^{(1)})| \geq \cdots \geq |t(x^{(N!)})|$. Let k be the largest integer $\leq N! \alpha + 1$. Then the hypothesis that the two samples came from the same (arbitrary) population is rejected if and only if $|t(x)| > |t(x^{(k)})|$, and the size of the test is $\leq \alpha$. The critical value $|t(x^{(k)})| = \lambda_N'$, say, is a random variable. It is shown that as $N \to \infty$, λ_N' tends in probability to λ under general conditions which cover the case of two samples from two normal populations. It follows that in large samples the power of the nonparametric test approaches that of the standard parametric test. Similar results hold for tests of certain linear hypotheses, the correlation coefficient test, etc. (Work sponsored by the Office of Naval Research.)

13. A Complete Class of Decision Procedures for Distributions with Monotone Likelihood Ratio. Herman Rubin, Stanford University.

Let $P(X \leq x \mid \tau) = \int_{-\infty}^{x} f(x, \tau)\, d\mu(x)$, where τ lies in some interval of the reals, and if $x_1 > x_2$, $\tau_1 > \tau_2$, then $f(x_1, \tau_1)f(x_2, \tau_2) - f(x_2, \tau_1)f(x_1, \tau_2) \geq 0$. This is a generalization of the exponential family, where $f(x, \tau) = \omega(\tau) \exp(x, \tau)$. Suppose the terminal decision d ranges over a closed subset D of the reals. Then if the loss function $W(d, \tau)$ satisfies certain monotonicity restrictions (which are usually met in multiple decision and estimation problems), a complete class of decision procedures based on a single observation are those which are unrandomized, except possibly at jumps of μ, and are monotone.

14. Some Nonparametric Results for Experimental Designs. John E. Walsh, Bureau of the Census.

In experimental designs, the quantities investigated are often grouped into blocks as a method of obtaining a higher precision for the experiment. This grouping may result in high correlation among observations within the same block. Also there may be substantial variance differences between blocks. Then the t-statistic is not necessarily applicable for comparing the effects of the treatments under investigation. This paper presents some nonparametric results which are usually valid for a well known type of experimental design if there is statistical independence among blocks (number of blocks ≥ 4). These nonparametric results are reasonably efficient, compared to those based on the t-statistic, for the case where the totality of observations are independent, normally distributed, and have the same variance. High precision can sometimes be obtained by designing the experiment to yield large positive correlation within blocks and then using the nonparametric results.

15. Efficient Tests and Confidence Intervals for Mortality Rates. John E. Walsh, Bureau of the Census.

This paper presents large-sample tests and confidence intervals for the "true value" of a mortality rate based on insurance data. These results have efficiencies of nearly 100%; i.e., they utilize nearly all the "information" contained in the data. The procedure used in obtaining the tests and confidence intervals consists in constructing a suitable t-statistic. The construction requires that the data be divided into between 300 and 400 statistically independent subgroups of approximately the same size. One possible way of accomplishing this is by subdividing the data according to the first three letters of the last name of the person insured and then appropriately combining the resulting groups. The amount of work required in applying the results of this paper is not appreciably greater than that

required in obtaining the usual point estimate of the mortality rate; in fact, a procedure which yields the point estimate as a byproduct is followed.

16. **Sufficient Statistics when the Carrier of the Distribution Depends on the Parameter.** D. A. S. FRASER, University of Toronto.

A "statistic of selection" is defined by a mapping from the space of the distribution to the space of Borel sets over that space. This statistic is sufficient if the parameter is a "parameter of selection," that is, if the parameter θ determines only the carrier of the distribution, the relative density being independent of the parameter. For more general distributions a theorem in this paper facilitates obtaining sufficient statistics, subject to continuity conditions.

17. **Bayes Solutions and Likelihood Ratio Tests of Some Simple and Composite Hypotheses. (Preliminary Report.)** ALLAN BIRNBAUM, Columbia University.

Let H_0 be a hypothesis concerning the density function $p_\theta(e)$, to be tested against the composite alternative H_1, by means of the acceptance region A in the space of the minimal sufficient statistic $t(e)$. For various distributions in which $t(e)$ is not real-valued, necessary and sufficient conditions are given for A to be a Bayes solution or the limit of a sequence of Bayes solutions. The likelihood ratio test, for a wide class of simple and composite hypotheses, is proved to be the limit of a sequence of Bayes solutions. A condition which is necessary and sufficient for the admissibility of the likelihood ratio test is derived. The distributions considered include: (1) $p_\theta(e)$ of general Koopman-Darmois form; the result here is applied to various examples. (2) $p_\theta(e)$ rectangular; generalizations of this result are indicated. Methods of approximating these tests are discussed. Applications to problems of "combining" independent significance tests are made; a definition of admissibility of methods of combination is proposed, according to which some current methods are inadmissible; a minimax multidecision procedure is proposed and developed, to replace certain current methods of combining tests.

18. **The Impossibility of Certain Affine Resolvable Balanced Incomplete Block Designs.** S. S. SHRIKHANDE, Nagpur College of Science, India.

Three theorems on the impossibility of an Affine Resolvable Design (R. C. BOSE, "A note on the resolvability of balanced incomplete designs," *Sankyhā*, Vol. 6 (1942), pp. 105-110) with parameters $v = nk = n^2(n-1)t + n^2$, $b = nr = n(n^2t + n + 1)$, $\lambda = nt + 1$, with $n \geq 2, t \geq 0$ (n and t integral) are proved. THEOREM 1. *An Affine Resolvable Design with the above parameters does not exist when n and t are odd and* (i) $n((n-1)t+1)$ *is not a perfect square, or* (ii) $n((n-1)t+1)$ *is a perfect square and* $nt + 1 \equiv 2 \pmod 4$, *and the square-free part of n contains a prime of the form* $4i + 3$. THEOREM 2. *An Affine Resolvable Design with the above parameters does not exist when n is odd and t is even and* (i) $(n-1)t + 1$ *is not a perfect square, or* (ii) $(n-1)t + 1$ *is a perfect square and* $n + t \equiv 1 \pmod 4$, *and the square-free part of n contains a prime* $4i + 3$. THEOREM 3. *An Affine Resolvable Design with the above parameters does not exist for any value of t if* $n \equiv 2 \pmod 4$ *and the square-free part of n contains a prime* $4i + 3$. The proofs depend on showing the impossibility of a Group Design obtained from the Affine Resolvable Design by making use of results due to Bose and Connor (Abstracts No. 4 and No. 6, *Annals of Math. Stat.*, Vol. 22 (1951), pp. 311-312). The theorem on the impossibility of finite projective planes (R. H. BRUCK AND H. J. RYSER, "The nonexistence of certain finite projective planes," *Canadian Jour. Math.*, Vol. 1 (1949), pp. 88-93) is contained here as a particular case.

19. **On Sufficiency and Statistical Decision Functions.** RAGHU RAJ BAHADUR, University of Chicago and Delhi University, India.

The first part of the paper contains certain characterizations of sufficiency. These results are then used to show that the justification for the use of sufficient statistics in statistical methodology which was described in an informal way by P. R. Halmos and L. J. Savage in the final section of their work on sufficiency ("Application of the Radon-Nikodym theorem to the theory of sufficient statistics," *Annals. of Math. Stat.*, Vol. 20 (1949), pp. 225–241) is valid whenever the decision space may be taken to be a subset of Euclidean k-space. This justification is proved first for the case of an arbitrary but fixed sample space, and then generalized to sequential sample spaces. The result for the sequential case may be outlined as follows. Let $x_1, x_2, \cdots$ be a sequence of chance quantities having a joint probability distribution p belonging to a family P. For each $m = 1, 2, \cdots$, let $t_m(x_1, x_2, \cdots, x_m)$ be a statistic which is sufficient when the sample space consists of points $(x_1, x_2 \cdots, x_m)$. Then corresponding to any sequential decision function ξ based on $x_1, x_2, \cdots$ there exists a sequential decision function η based on $t_1(x_1)$, $t_2(x_1, x_2)$, $\cdots$ such that the joint probability distribution of the sample size and the terminal decision is the same under ξ and η for each p in P. This result holds without restriction (other than measurability) on the sampling scheme of ξ, so that in the special case of point estimation with a convex loss function it leads to an enlargement of the domain of Blackwell's Theorem and its generalizations.

20. **A Two Sample Test Procedure.** DONALD B. OWEN, University of Washington.

In testing hypotheses the standard procedure is to specify a test based on a single set of observations. Sequential analysis introduced a new concept: that of making a decision after each observation, either to accept the hypothesis, or to reject it, or to take another observation. Here an approach is worked out that lies somewhat between these: an initial set of observations is taken. Then a decision is made to accept, reject, or take one more set of observations. After this second set of observations, a decision on the hypothesis must be made.

The problem is to formulate these decision rules at the two stages of the process, to optimize them if possible, and to evaluate the performance of the tests. These depend on various parameters and it is pertinent to inquire how these parameters affect the answers to the questions noted. More precisely, the problem is to maximize (or minimize) with respect to a, b, c, d expressions of the type: $\int_a^b f(x)\,dx$, $\int_c^d g(x)\,dx$, subject to side conditions (which are also expressed as integrals). The functions f and g are probability density functions: here those associated with the normal probability density function. There are two main sections, the basis of division being whether the final decision is based only on the second sample or on the whole set of observations. The second procedure is the more economical, but mathematically it is much more difficult. Much less complete results are given in this section. In the direction of finding the optimum of this type of rule, the results are chiefly negative. Several theorems are given which show the difficulties in obtaining a solution to this problem For the rules given the performance of the tests is evaluated and various theorems worked out concerning them. Some of the lemmas have interest in their own right as properties pertaining to important probability density functions.

21. **A Combinatorial Central Limit Theorem.** WASSILY HOEFFDING, University of North Carolina.

Let $(Y_{n1}, \cdots, Y_{nn})$ be a random vector which takes on the $n!$ permutations of $(1, \cdots, n)$ with equal probabilities. Let $c_n(i, j)$, $i, j = 1, \cdots, n$, be n^2 real numbers. Two sufficient conditions for the asymptotic normality of $S_n = \sum_{i=1}^n c_n(i, Y_{ni})$ are given. In

the special case $c_n(i, j) = a_n(i)b_n(j)$, which was considered by Wald and Wolfowitz, the first condition generalizes a condition given by Noether ("On a theorem by Wald and Wolfowitz," *Annals of Math. Stat.*, Vol. 20 (1949), pp. 455–458). The second condition is slightly stronger but simpler as it involves not an infinity of limiting relations but only a single one. Applications to the theory of nonparametric tests are indicated. (Work sponsored by the Office of Naval Research.)

22. Necessary Conditions for the Existence of a Symmetrical Group Divisible Design. R. C. Bose and W. S. Connor, Jr., University of North Carolina.

An incomplete block design with v treatments each replicated r times in b blocks of size k is said to be group divisible if the treatments can be divided into m groups each with n treatments, so that the treatments of the same group occur together in λ_1 blocks, and treatments of different groups occur together in λ_2 blocks, $\lambda_1 \neq \lambda_2$. The combinatorial properties and the methods of construction for these designs have been studied by the authors elsewhere (cf. Abstract No. 4, *Annals of Math. Stat.*, Vol. 22 (1951), p. 311). An incomplete block design is said to be symmetrical if the number of treatments v equals the number of blocks b, and in consequence $k = r$. If N is the incidence matrix of a symmetrical group divisible design, the Hasse invariate $C_p(NN')$ of the quadratic form with matrix NN' (where N' is the transpose of N, and p is any odd prime) has been obtained in a simple form. Its value is $C_p(NN') = (P, \lambda_2)_p(P, n)_p^m(P, -1)_p^{m(m-1)/2}(Q, n)_p^m(Q, -1)_p^{v(v-1)/2+m(m-1)/2}$, where $P = r^2 - v\lambda_2$, $Q = r - \lambda_1$, and $(a, b)_p$ is the Hilbert norm residue symbol. For the existence of a symmetrical group divisible design $C_p(NN') = +1$ for all odd primes p, and $P^{m-1}Q^{m(n-1)}$ must be a perfect square. This shows that necessary conditions for the existence of a symmetrical group divisible design are (i) if m is even then P must be a perfect square, and if further $m = 4t + 2$ and n is even then $(Q, -1)_p = +1$ for any odd prime p; (ii) if m is odd and n is even then Q must be a perfect square and $((-1)^{\alpha}n\lambda_2, P)_p = +1$ for any odd prime p, where $\alpha = m(m-1)/2$; (iii) if m and n are both odd then $((-1)^{\alpha}n\lambda_2, P)_p((-1)^{\beta}n, Q)_p = +1$ for any odd prime p, where $\alpha = m(m-1)/2$ and $\beta = n(n-1)/2$. The impossibility of a large number of symmetrical group divisible designs can be proved by using these conditions.

23. On a Problem of Mapping of One Space on Another with Applications in Sampling Distributions. S. N. Roy, University of North Carolina.

Denoting a $p \times n$ matrix M by $M(:p \times n)$, a triangular matrix (with the upper right hand corner zero) by $\tilde{T}$ (with elements t_{ij}), a diagonal matrix with diagonal elements $\theta_1, \theta_2, \cdots, \theta_p$ by D_θ, and a $p \times p$ unit matrix by $I(p)$, consider the transformations

(i) $x(:p \times n) = \tilde{T}(:p \times p)\, \mathcal{L}\, (:p \times n)$, where $p \leq n$; x is of rank p; $\mathcal{L}\mathcal{L}' = I(p)$; $t_{ii} > 0$, $i = 1, 2, \cdots, p$.

(ii) $x(:p \times n) = M(:p \times p)D_{\sqrt{\theta}}(:p \times p)\mathcal{L}(:p \times n)$, where $MM' = M'M = \mathcal{L}\mathcal{L}' = I(p)$; $p \leq n$; x is of rank p; the first row of M consists of positive elements; θ stands for the p positive characteristic roots of xx', and $\sqrt{\theta}$ stands for the positive square root of θ.

These transformations have proved extremely useful for almost the entire range of problems on sampling distributions based on multivariate normal populations. In (i), by virtue of $\mathcal{L}\mathcal{L}' = I(p)$, we could choose from $\mathcal{L}$, in various alternative ways, a set of $pn - p(p + 1)/2$ independent elements to be called $\mathcal{L}_I$, and in (ii), by virtue of $MM' = I(p)$, a similar set of $p(p - 1)/2$ from M to be called M_I, and by virtue of $\mathcal{L}\mathcal{L}' = I(p)$ a similar set of $pn - p(p + 1)/2$ from $\mathcal{L}$ to be called $\mathcal{L}_I$. In this paper is discussed the nature of the transformations (i*a*), from x to $\tilde{T}\mathcal{L}$, under $\mathcal{L}\mathcal{L}' = I(p)$; (i*b*) from x to $\tilde{T}\mathcal{L}_I$; (ii*a*) from x to $\theta M\mathcal{L}$, under $MM' = M'M = \mathcal{L}\mathcal{L}' = I(p)$; and (ii*b*) from x to $\theta M_I\mathcal{L}_I$. In this con-

nection certain problems are also posed for mathematical statisticians which the author has not been able to solve so far.

24. On a Theorem in Jacobians with Statistical Applications. S. N. Roy, University of North Carolina.

If $F_i(y_1, \cdots, y_m, x_1, \cdots, x_{m+n}) = 0$ $(i = 1, 2, \cdots, m + n)$ are such that we could select any set of m x_i's (to be called without any loss of generality $x_1, x_2, \cdots, x_m$) and could find real values of $(y_1, \cdots, y_m)$ and of $(x_{m+1}, \cdots, x_{m+n})$ for real values of $(x_1, \cdots, x_m)$, then, assuming that the numerator and the denominator on the right-hand side of the equation below are nonvanishing, and assuming certain other restrictions, we would have $J(y_1, y_2, \cdots, y_m : x_1, x_2, \cdots, x_m) = -[\partial(F_1, \cdots, F_{m+n})/\partial(x_1, \cdots, x_{m+n})]/[\partial(F_1, \cdots, F_{m+n})/\partial(y_1, \cdots, y_m, x_{m+1}, \cdots, x_{m+n})]$, where absolute values of the determinants are to be taken. Important special cases of this general theorem with various statistical applications are discussed in this paper.

25. The Inventory Problem. A. Dvoretzky, J. Kiefer, and J. Wolfowitz, Cornell University.

The inventory problem is the general problem of what quantities of goods to stock in anticipation of future demand, where loss is caused by inability to supply demand or by stocking goods for which there is no demand. Let x_i be the initial stock of a given commodity in the ith interval $(i = 1, \cdots, N)$ before any ordering is done, and y_i the starting stock after an amount $y_i - x_i \geq 0$ has been ordered and instantaneously received by the stocking agency. The amount demanded in the ith interval is a chance variable with known distribution function F_i. $W_i(x_i, y_i, d_i)$ is the loss incurred in the ith interval when x_i is the starting stock, y_i the initial stock, and d_i is the amount demanded in this interval. F_i, W_i, and the expected value of W_i with respect to the demand may also be functions of the "past history" as given by $\beta_i = \{x_j, y_j, d_j : j < i\}$. An ordering policy Y is a set of functions $Y_i(x_i, \beta_i)$ $(i = 1, \cdots, N)$, where one orders an amount $Y_i(x_i, \beta_i) - x_i$ in the ith interval. With each Y and x_1 there is associated a quantity $A(Y \mid x_1)$ which is the total expected loss over all intervals (the loss in the ith interval being discounted by a factor $1 - \alpha_i$) when Y is used and x_1 is the initial stock in the first interval. An optimal (ϵ-optimal) policy is one which minimizes this quantity (within ϵ) for every x_1. A method for constructing such policies is given. The case of an infinite number of intervals is similarly treated. Analogous results are obtained in more general cases, e.g., when there is a time lag between the ordering and delivery of goods, when there are several commodities, etc.

The second part of the paper deals with the case when the set of distributions F_i is known only to be a member of a certain class Ω. Constructive methods for obtaining Bayes solutions and complete classes are given. (This research was sponsored by the Office of Naval Research.)

NEWS AND NOTICES

Readers are invited to submit to the Secretary of the Institute news items of interest

Personal Items

Mr. Fred C. Andrews, Teaching Assistant and Research Assistant, Statistical Laboratory, University of California, Berkeley, was promoted to Lecturer and Research Assistant effective July 1, 1951.

Dr. Dorothy S. Brady, formerly Professor of Economics at the University of

Illinois, has returned to Washington as Consultant to the Commissioner on Costs and Standards of Living, Bureau of Labor Statistics.

Dr. C. West Churchman has taken a leave of absence for one year from Wayne University to accept a visiting professorship in the Engineering Administration Department, Case Institute of Technology, Cleveland, Ohio, to do work in operations research as applied to industry.

Mr. Edward L. Corton, Jr., formerly employed at the Naval Ordnance Training Station, China Lake, California, is now working as a meteorologist for the Navy Hydrographic Office, Washington, D. C.

Dr. Arthur M. Dutton, after receiving his degree in June from the Iowa State College, accepted a position with Professor S. L. Crump at the University of Rochester Atomic Energy Project, Rochester, New York.

Dr. Evelyn Fix, Instructor and Research Associate, Statistical Laboratory, University of California, Berkeley, was promoted to Assistant Professor and Research Associate, effective July 1, 1951.

Mr. Charles P. Gershenson accepted the position as Research Director of the Jewish Children's Bureau of Chicago after working for five and one-half years as Research Associate at the Institute of Psychological Research, Teachers College, Columbia University. His new job entails the development of a research program for evaluating the effectiveness of small residential units for the treatment of emotional disturbed children.

Dr. E. J. Gumbel has been appointed Consultant to the Applied Mathematics and Statistics Laboratory, Stanford University, Stanford, California, for work on industrial statistics.

Dr. Louis Guttman, formerly Professor of Sociology at Cornell University, has accepted a position with the Israel Institute of Applied Social Research, Shell Building, Julian's Way, Jerusalem.

Dr. Paul R. Halmos will be on leave of absence from the University of Chicago for the academic year 1951–1952. He will serve as Visiting Professor of Mathematics at the University of Montevideo under the auspices of the State Department's Division of International Exchange of Persons.

Dr. J. L. Hodges, Jr., Assistant Professor and Research Associate, Statistical Laboratory, University of California, Berkeley, will be on leave of absence in a visiting capacity at the University of Chicago during the academic year 1951–1952.

Mr. Daniel G. Horvitz has accepted an assistant professorship of Biostatistics at the University of Pittsburgh School of Public Health.

Mr. William C. James has left Washington to serve with the National Office of Vital Statistics, Public Health Service, in Lima, Peru. Last fall he was appointed as international consultant in vital statistics under the international vital statistics cooperation program of this department and was in Washington for several months training for this work.

Mr. T. A. Jeeves, Lecturer and Research Associate, Statistical Laboratory, University of California, Berkeley, has been promoted to Instructor and Research Associate, effective July 1, 1951.

Mr. Jack Kiefer will be an Instructor in Mathematics at Cornell University during 1951–1952.

Dr. Erich L. Lehmann, Assistant Professor and Research Associate, Statistical Laboratory, University of California, Berkeley, has been promoted to Associate Professor and Research Associate, effective July 1, 1951. During the academic year 1951–1952 Professor Lehmann will be in a visiting capacity at Stanford University, on leave from the University of California.

Mr. Edward M. Schrock, formerly Division Engineer, General Electric Co., Erie, Pennsylvania, is now Supervisor of Quality Control, Lukens Steel Company, Coatesville, Pennsylvania.

Dr. Elizabeth L. Scott, Instructor and Research Associate, Statistical Laboratory, University of California, Berkeley, was promoted to Assistant Professor and Research Associate, effective July 1, 1951.

Dr. Irving H. Siegel has left Johns Hopkins University, where he served as Lecturer in Political Economy and as Director of Productivity Studies in the Operations Research Office, to become Director of the American Technology Study for the Twentieth Century Fund in Washington.

Dr. Andrew Sobczyk is on leave of absence from Boston University to serve as a Staff Member at Los Alamos Scientific Laboratory, University of California. Los Alamos, New Mexico.

Mr. Robert F. Tate, Associate and Research Assistant, Statistical Laboratory, University of California, Berkeley, was promoted to Lecturer and Research Assistant, effective July 1, 1951.

Dr. Shanti A. Vora has resigned his assistant professorship in the Department of Statistics, Stanford University, to accept a position as Statistician with the Standard-Vacuum Oil Co., India Division Office, Bombay.

Awards for Post-doctoral Study in Statistics at the University of Chicago

The Committee on Statistics (a department) of the University of Chicago has established, under a five-year grant from the Rockefeller Foundation, a program of Post-doctoral Awards to provide training and experience in statistics for scholars whose main interests lie outside that field. There will be three Awards per year, to holders of the doctorate or equivalent in the biological, the physical, and the social sciences. Each Award will be $4000 or slightly more, office space will be provided, and $600 to $1000 will be available for clerical, computational, and research assistance. There will be no tuition charges.

The purpose of the Awards is to give statistical training to a few scientists who may be expected to employ it both to the direct advance of their specialties and to the enlightenment of their colleagues and students by example, by consultation, and by formal instruction. The development of the field of statistics has been so rapid that problems of communication are a serious obstacle to its full exploitation. The amount and quality of instruction available to young students is constantly increasing, but there is a real need, which these Awards seek to fill, for making appropriate instruction available to already established

scientific workers who give promise of immediate applications of statistics to their special fields.

Recipients of the Awards must have received the doctor's degree prior to commencing the program, except in the case of recognized research workers whose experience and accomplishments are clearly the equivalent. Candidates whose mathematical preparation includes less than the usual sophomore year of calculus, or its equivalent, will not ordinarily be considered, but previous training in statistics is not required or expected. Candidates having under way research programs in their own fields will be preferred, and the department of the University of Chicago concerned with a candidate's specialty will be asked to participate in evaluating his application. Recipients must spend eleven months studying statistics at the University of Chicago, and will be expected to pursue a number of regular courses.

Applications, or requests for further information, should be sent to: Committee on Statistics, University of Chicago, Chicago 37. Applications for the academic year 1952–53 should arrive by April 1, 1952.

Proceedings of the Second Berkeley Symposium

The Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, 1950, contains forty-six articles on mathematical statistics, probability, and their applications to astronomy, biometry, econometrics, physics, traffic engineering, and wave analysis, by the following authors: T. W. Anderson, K. J. Arrow, E. W. Barankin, D. M. Belmont, J. Berkson, D. S. Berry, D. Blackwell, G. W. Brown, K. L. Chung, W. G. Cochran, H. Cramér, B. de Finetti, J. L. Doob, A. Dvoretzky, P. Erdös, W. Feller, R. P. Feynman, T. W. Forbes, R. Fortet, M. A. Girshick, T. E. Harris, L. G. Henyey, J. L. Hodges, Jr., W. Hoeffding, P. G. Hoel, H. Hotelling, M. Kac, S. Kakutani, J. Kampé de Fériet, H. W. Kuhn, R. S. Lehman, E. L. Lehmann, V. F. Lenzen, P. Lévy, H. W. Lewis, B. Lindblad, M. Loève, J. Marschak, A. M. Mood, G. Placzek, H. Robbins, P. Rudnick, L. J. Savage, E. L. Scott, H. R. Seiwell, O. Struve, R. J. Trumpler, A. W. Tucker, A. Wald, W. A. Wallis, J. Wolfowitz, A. Zygmund.

The price of this volume of 666 pages is $11.00. Orders for the Proceedings should be addressed to University of California Press, Berkeley 4, California.

New Members

The following persons have been elected to membership in the Institute

(June 1, 1951 to August 22, 1951)

Allen, Stephen G., Jr., M.A. (Univ. of Chicago), Research Associate, Applied Mathematics and Statistics Laboratory, Stanford University, Stanford, California.

Amundsen, Mrs. Herdis T., Aktuarkandidat (Univ. of Oslo), Lecturer in Mathematical Statistics, University of Oslo, Oslo, Norway.

Appel, Frederick W., Ph.D. (Univ. of Chicago), Executive Secretary, Public Health and Experimental Therapeutics Study Sections, Division of Research Grants, National Institutes of Health, Public Health Service, *3450 – 38th Street, N.W., Washington 16, D. C.*

Berg, William D., Ph.D. (Univ. of Iowa), Assistant Professor, University of Ohio, *Box 285, Gambier, Ohio.*

Botzum, Rev. William A., Ph.D. (Univ. of Chicago), Instructor, Department of Education, University of Notre Dame, Notre Dame, Indiana.

Davison, George H., Secretary, The United Steel Companies, Ltd., Research & Development Department, Swinden House, Moorgate, Rotherham, England.

de Agrisqueta, Francisco, Ph.D. (Univ. de Deusto, Bilbao, Spain), Acting Secretary General of Inter American Statistical Institute, *1801 Clydesdale Place, N.W., Apartment 209, Washington 9, D. C.*

de la Garza, Andres, B.S., Statistician, *105 Pacific, Oak Ridge, Tennessee.*

Foscue, Augustus W., Jr., M.B.A. (Stanford Univ.), Professor of Accounting and Statistics and Chairman of Statistics Department, School of Business Administration, Southern Methodist University, Dallas 5, Texas.

Fritz, Edward, M.S. (Univ. of Michigan), Research Engineer, Franklin Institute, *4025 Girard Avenue, Philadelphia, Pennsylvania.*

Goulden, Cyril H., Ph.D. (Univ. of Minnesota), Chief, Cereal Division, Experimental Farm, Ottawa, Canada.

Grant, J. Douglas, M.A. (Stanford Univ.), Clinical Psychologist, U. S. Naval Retraining Command, Mare Island, Vallejo, California.

Green, Bert F., Jr., Ph.D. (Princeton), Staff Member, Research Laboratory of Electronics, Room 22A—233c, Massachusetts Institute of Technology, Cambridge 39, Massachusetts.

Hadden, Stuart T., M.A. (Temple Univ.), Chemical Engineer and Senior Technologist, Research and Development Department, Socony-Vacuum Oil Co., Inc., Paulsboro, New Jersey, *416 S. Jackson Street, Woodbury, New Jersey.*

Hildreth, Clifford, Ph.D. (Iowa State College), Associate Professor, Cowles Commission for Research in Economics, University of Chicago, Chicago 37, Illinois.

Howe, William G., B.A. (Univ. of Rochester), Statistician, Color Control Department, Kodak Park, Eastman Kodak Company, Rochester, New York.

Jackson, Patricia L., Ph.D. (Columbia Univ.), Employment Manager, Alexander's Department Stores, Inc., Grand Concourse & Fordham Road, Bronx, New York.

Kaelin, Alois, Dipl. Math. (Eidgenossischen Technischen Hochschule, Zurich), *Waldlistr. 5, Zurich 7/32, Switzerland.*

Kibbey, Milton E., B.A. (Univ. of Michigan), Development Engineer and Member of Statistical Analysis Group, *351 W. Outer Drive, Oak Ridge, Tennessee.*

Kuffner, Peter K., B.S. (Univ. of Chicago), Student, Department of Mathematics, University of Chicago, *4002 S. Brighton Place, Chicago 32, Illinois.*

Kurkjian, Badrig M., S.B. (Mass. Inst. of Tech.), Mathematician, Room 300, Building 92, National Bureau of Standards, *217 Webster Street, N.E., Washington, D. C.*

Lewsley, Bernard J., B.S. (Birmingham Univ., England), Research and Development Engineer, General Electric Company, Ltd., *72 Ansty Road, Coventry, England.*

Marans, Frances A., B.S. (Wilson Teachers College, Wash., D. C.), Mathematical Statistician, Office of the Statistical Consultant, Division of Manpower and Employment Statistics, Bureau of Labor Statistics, Department of Labor, *1336 Missouri Avenue, N.W., Apt. 218, Washington 11, D. C.*

Mitten, Loring G., M.S. (Mass. Inst. of Tech.), Instructor, Department of Industrial Engineering, Ohio State University, Industrial Engineering Building, Columbus 10, Ohio.

Moonan, William J., M.A. (Univ. of Minnesota), Instructor in Statistics Laboratory, College of Education, University of Minnesota, *23 E. 54th Street, Minneapolis, Minnesota.*

Moore, Cordell B., M.S. (Univ. of Kentucky), Instructor, Department of Mathematics, University of Kentucky, Lexington, Kentucky.

Sachs, David, A.B. (Columbia Univ.), Student, Department of Mathematical Statistics, Columbia University, *75 Central Park West, New York 23, New York.*

Smith, Cecil W., Analysis Officer, No. 3 Line, British Overseas Airways Corporation, *107 Pembroke Road, Clifton, Bristol 8, England.*

Stalnaker, John M., M.A. (Univ. of Chicago), Director of Studies, Association of American Medical Colleges, *1075 Elm Street, Winnetka, Illinois.*

Swalm, R. O., B.S. (Univ. of Pa.), Assistant Professor of Industrial Engineering, Syracuse University, *9 Wyncrest Drive, East Syracuse, New York.*

Thomson, Kenneth F., Ph.D. (Ohio State Univ.), Assistant Project Director, Richardson, Bellows, Henry & Co., 439 Madison Avenue, New York City, *25 Burbank Street, Yonkers 2, New York.*

Titman, Richard H., B.S. (Univ. of Wash.), Statistician, General Electric Co., Nucleonics Division, *711 Stanton, Richland, Washington.*

Vinci, Felice, Doctor Juris. (Univ. of Palermo, Italy), Professor of Statistics and Director of Instituto di Scienze Economiche e Statistiche, University of Milan, *Via Lamarmora 42, Milano, Italy.*

White, Aubrey, B.A. (Univ. of Toronto), Consulting Actuary and member of Ostheimer & Co., 1500 Chestnut Street, Philadelphia, Pennsylvania.

White, John S., M.A. (Univ. of Minnesota), Graduate Student, University of Minnesota, *134 West 62nd Street, Minneapolis 19, Minnesota.*

Whitney, Alfred G., Ed.M. (Harvard), Research Associate, Life Insurance Agency Management Association, *855 Asylum Avenue, Hartford 5, Connecticut.*

Winer, Ben J., M.S. (Univ. of Oregon), Research Associate, Personnel Research Board, Ohio State University, Columbus, Ohio.

Wittenborn, J. R., Ph.D. (Univ. of Illinois), Research Associate in Psychology, Yale University, *33 Cedar Street, New Haven, Connecticut.*

Zobel, Sigmund P., M.B.A. (Univ. of Buffalo), Lecturer in Statistics, School of Business Administration, also Assistant in Preventive Medicine and Public Health, School of Medicine, University of Buffalo, *105 Landon Street, Buffalo 8, New York.*

REPORT OF THE MINNEAPOLIS MEETING OF THE INSTITUTE

The thirteenth summer meeting and forty-eighth meeting of the Institute of Mathematical Statistics was held at the University of Minnesota, September 4–7, 1951, in conjunction with the summer meeting of the Mathematical Association of America, the summer meeting of the American Mathematical Society and the Minneapolis meeting of the Econometric Society. The meeting was attended by the following one hundred and six members of the Institute:

G. E. Albert, R. L. Anderson, T. W. Anderson, K. J. Arnold, W. D. Baten, Helen P. Beard, Agnes Berger, Joseph Berkson, Jean Bronfenbrenner, Irwin Bross, Hobart Bushey, Maria Castellani, Herman Chernoff, A. G. Clark, T. F. Cope, A. H. Copeland, Sr., L. M. Court, E. L. Cox, J. F. Daly, G. B. Dantzig, W. J. Dixon, J. L. Doob, Aryeh Dvoretzky, P. S. Dwyer, Churchill Eisenhart, Lillian Elveback, H. P. Evans, C. H. Fischer, Evelyn Fix, J. S. Frame, Robert Gage, H. M. Gehman, L. A. Goodman, John Gurland, P. C. Hammer, T. E. Harris, H. L. Harter, Clifford Hildreth, J. L. Hodges, Jr., Wassily Hoeffding, J. F. Hofmann, Robert Hogg, Harold Hotelling, H. M. Hughes, S. L. Isaacson, P. O. Johnson, Walbert Kalinowski, Leo Katz, Harriet J. Kelly, O. Kempthorne, M. G. Kendall, Jack Kiefer, W. M. Kincaid, T. C. Koopmans, C. F. Kossack, R. L. Kozelka, William Kruskal,

O. E. Lancaster, F. C. Leone, Howard Levene, S. B. Littauer, R. B. McHugh, H. B. Mann, Margaret P. Martin, K. O. May, G. F. T. Mayer, Paul Meier, M. R. Mickey, Sigeiti Moriguti, Frederick Mosteller, Jerzy Neyman, M. L. Norden, I. Olkin, Arthur Ollivier, Toby Oxtoby, M. P. Peisakoff, G. B. Price, J. A. Rafferty, Howard Raiffa, F. D. Rigby, D. D. Rippe, Herman Rubin, Henry Scheffé, Elizabeth L. Scott, W. B. Simpson, Andrew Sobczyk, Milton Sobel, M. D. Springer, Charles Stein, Joseph Talacko, W. F. Taylor, Henry Teicher, D. Teichroew, D. J. Thompson, L. J. Tick, G. Tintner, A. E. Treloar, A. W. Tucker, J. W. Tukey, S. A. Tyler, E. C. Varnum, John von Neumann, D. F. Votaw, J. S. White, J. Wolfowitz, and M. A. Woodbury.

The meeting opened on Tuesday morning, September 4, 1951, with a *Symposium on Medical Statistics*. Professor Alan Treloar of the University of Minnesota was chairman. Dr. Joseph Berkson of the Mayo Clinic presented a paper entitled *Estimate of Effectiveness of Cancer Therapy from Mortality following Treatment*. Professor Jerzy Neyman of the University of California presented a paper entitled *Further Results Concerning the Follow-up Procedure*. Prepared discussion of the papers was offered by Miss Lillian Elveback of the University of Minnesota, Dr. Evelyn Fix of the University of California and Dr. W. F. Taylor of the School of Aviation Medicine. Approximately 55 persons attended the session.

During the meeting the Institute joined the Econometric Society in three sessions devoted to a *Symposium on the Theory of Games, Decision Problems and Related Topics*. The first of these sessions with the title *Theory of Games* for the session was held on Tuesday afternoon under the chairmanship of Professor John von Neumann of the Institute for Advanced Study. Over 250 persons attended this session. Papers were presented by Professor Samuel Karlin of Princeton University and the Rand Corporation and by Dr. Olaf Helmer of the Rand Corporation. Prepared discussion was presented by Dr. Seymour Sherman of Lockheed Aircraft Corporation, Dr. L. S. Shapley of Princeton University and the Rand Corporation, Professor Howard Raiffa of the University of Michigan, and Dr. G. B. Dantzig of the Department of the Air Force.

On Wednesday morning the first of three sessions for contributed papers was held under the chairmanship of Professor W. J. Dixon of the University of Oregon. Attendance was approximately 50. The following papers were presented:

1. *On Stieltjes Integral Equations of Stochastic Processes*. Maria Castellani, University of Kansas City.
2. *An Unfavorable Aspect of the Likelihood Ratio Test*. L. M. Court, Rutgers University.
3. *Impartial Decision Rules and Sufficient Statistics*. R. R. Bahadur and L. A. Goodman, University of Chicago.
4. *Contributions to the Statistical Theory of Counter Data*. G. E. Albert, University of Tennessee and Oak Ridge National Laboratory, and M. L. Nelson, Oak Ridge National Laboratory.
5. *On the Use of Wald's Classification Statistic*. H. L. Harter, Michigan State College.

Later on Wednesday morning, Session II—*Statistical Decision Problems*—of the *Symposium on the Theory of Games, Decision Problems and Related Topics* was held with Professor A. W. Tucker of Princeton University as chairman. Approximately 175 persons attended. Papers were presented by Professor David

Blackwell of Howard University and Professor Charles Stein of the University of Chicago. Prepared discussion was presented by Dr. M. P. Peisakoff of the Rand Corporation, Professor Aryeh Dvoretsky of Hebrew University, Jerusalem, and Cornell University, Professor Herman Chernoff of the University of Illinois, Professor J. L. Hodges of the University of California, and Professor Samuel Karlin of Princeton University and the Rand Corporation.

On Thursday morning a *Symposium on Probability and Statistical Inference* was held jointly with the Econometric Society. Professor Leo Katz of Michigan State College was chairman. Approximately 140 persons attended. Professor Jerzy Neyman of the University of California presented a paper on *Inductive Behavior* and Professor J. W. Tukey of Princeton University presented a paper on *Purposes of Fiducial Inference*. Prepared discussion was presented by Professor Leonid Hurwicz of the University of Minnesota and Professor Gerhard Tintner of Iowa State College.

The Third Rietz Memorial Lecture was given by Professor Harold Hotelling of the University of North Carolina at 2 p.m. Thursday, September 6, 1951 under the title, *The Behavior of Standard Statistical Tests under Nonstandard Conditions*. Professor Jerzy Neyman of the University of California was chairman of the session. The attendance was approximately 120.

Later on Thursday afternoon, Session III—*Decision Making and Theory of Organization*—of the *Symposium on the Theory of Games, Decision Problems and Related Topics* was held with Professor T. C. Koopmans of the Cowles Commission as chairman. Approximately 100 persons attended. A paper by Professor Leonid Hurwicz of the University of Minnesota was followed by prepared discussion by Dr. Norman Dalkey of the Rand Corporation, Professor David Gale of Brown University, and Commander W. H. Keen of the Naval Air Development Center. A paper by Dr. M. M. Flood of the Rand Corporation was followed by prepared discussion by Professor C. B. Tompkins of George Washington University and Professor H. A. Simon of Carnegie Institute of Technology.

Early Friday morning the second session for contributed papers, attended by approximately 50 people, was held with Professor Henry Scheffé of Columbia University as chairman. The following papers were presented:

6. *Polynomial Determination in a Field of Integers Modulo P*. E. C. Varnum, Barber-Colman Company.
7. *About Some Symmetrical Distributions from the Perks' Family of Functions*. J. V. Talacko, Marquette University.
8. *A Large-Sample Test for the Variation of Sample Covariance Matrices*. D. D. Rippe, University of Michigan.
9. *Probability Models for Analyzing Time Changes in Attitudes*. T. W. Anderson, Columbia University.
10. *The Variance of a Weighted Average Using Estimated Weights*. Paul Meier, Princeton University.

An *Abraham Wald Memorial Session* was held at 10:30 a.m. Friday, September 7, 1951. Professor Harold Hotelling of the University of North Carolina

was chairman. Approximately 100 people attended the session. A paper on *Wald's Contributions in Pure Mathematics* by Professor Karl Menger of Illinois Institute of Technology was read by Professor J. L. Kelley of Tulane University. A paper on *Wald's Contributions in Econometrics* was presented by Professor Gerhard Tintner of Iowa State College and a paper on *Wald's Contributions in Mathematical Statistics* was presented by Professor J. Wolfowitz of Cornell University.

Early Friday afternoon a session under the chairmanship of Professor Paul S. Dwyer of the University of Michigan was devoted to two invited addresses. Approximately 70 persons attended. Dr. T. E. Harris of the Rand Corporation spoke on *First Passage and Recurrence Distributions* and Professor M. G. Kendall of the London School of Economics spoke *On the Systematic Determination of Sampling Distributions.*

Later on Friday afternoon the third session for contributed papers was held under the chairmanship of Professor K. J. Arnold of the University of Wisconsin. Approximately 40 persons attended. The following papers were presented:

11. *Distribution of Ratios of Quadratic Forms.* John Gurland, University of Chicago.
12. *The Large-Sample Power of Tests Based on Permutations of Observations. Preliminary Report.* Wassily Hoeffding, University of North Carolina.
13. *A Complete Class of Decision Procedures for Distributions with Monotone Likelihood Ratio.* Herman Rubin, Stanford University.

The following papers were presented by title at the meeting:

14. *Some Nonparametric Results for Experimental Designs.* J. E. Walsh, Bureau of the Census.
15. *Efficient Tests and Confidence Intervals for Mortality Rates.* J. E. Walsh, Bureau of the Census.
16. *Sufficient Statistics when the Carrier of the Distribution Depends on the Parameter.* D. A. S. Fraser, University of Toronto.
17. *Bayes Solutions and Likelihood Ratio Tests of Some Simple and Composite Hypotheses. Preliminary Report.* Allan Birnbaum, Columbia University.
18. *The Impossibility of Certain Affine Resolvable Balanced Incomplete Block Designs.* S. S. Shrikhande, Nagpur College of Science, India.
19. *On Sufficiency and Statistical Decision Functions.* R. R. Bahadur, University of Chicago and Delhi University, India.
20. *A Two Sample Test Procedure.* D. B. Owen, University of Washington.
21. *A Combinatorial Central Limit Theorem.* Wassily Hoeffding, University of North Carolina.
22. *Necessary Conditions for the Existence of a Symmetrical Group Divisible Design.* R. C. Bose and W. S. Connor, Jr., University of North Carolina.
23. *On a Problem of Mapping of One Space on Another with Applications in Sampling Distributions.* S. N. Roy, University of North Carolina.
24. *On a Theorem in Jacobians with Statistical Applications.* S. N. Roy, University of North Carolina.
25. *The Inventory Problem.* A. Dvoretzky, J. Kiefer and J. Wolfowitz, Cornell University.

The Council of the Institute held a meeting at 1:30 p.m. on Wednesday, September 5, 1951.

A business meeting of the Institute was held at 9 a.m. on Thursday, September 6, 1951.

Social events included a reception sponsored by the mathematical organizations and the College of St. Thomas on Tuesday evening, a banquet on Wednesday evening, a piano recital on Thursday evening and a beer party sponsored by the Institute later on Thursday evening.

K. J. Arnold
Associate Secretary

PUBLICATIONS RECEIVED

Anuario Estadistico de España, (Instituto Nacional de Estadistica), Presidencia del Gobierno, Madrid, 1950, xliii + 1048 pp.

Anuario Estadistica de España (*Edicion Manual*), (Instituto Nacional de Estadistica), Presidencia del Gobierno, Madrid, 1951, lvii + 944 pp.

Walker, Helen M., *Mathematics Essential for Elementary Statistics*, rev. ed., Henry Holt and Company, New York, 1951, viii + 382 pp., \$2.75.

STATI[illegible] [illegible]TICS

President:
D. S. D[illegible]

President-Elect:
M. [illegible]

Secretary-Treasurer:
C. K. [illegible]

Editor:
T. W. An[illegible], Colum[illegible]

The purpose of [illegible] research in the mathematical [illegible] between the field of pure [illegible]

Membership dues including [illegible] are $10.00 per year [illegible] members. Dues and [illegible] the Institute [illegible] to the [illegible]

[illegible]

ANNUAL MEETING [illegible]